

COURSE PACK
FOR
Business Analytics
COURSE: BCA SEMESTER: VI
Course Code: 604

Year: 2021-22

Course Instructor: Ms.Tripti Tiwari *Tripti*

Course Leader: Ms. Tripti Tiwari *Tripti*

Checked by: Program Coordinator
Mr.Mahesh Chaubey

Forwarded by: HOD
Dr.Daljeet Singh Bawa

Yamini - Cg 1
Approved by: Director
Dr.Yamini Agarwal



Bharati Vidyapeeth Institute of Management & Research, New Delhi
An ISO 9001:2008 &14001:2004 Certified Institute
“A+” Grade Accredition by NAAC
(Ph: 01125284396, 25285808, Fax No. 01125286442)

FOR INTERNAL CIRCULATION ONLY



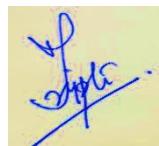
Checklist for Coursepacks

- Title page should be standardized bearing title of subject, course, course code, semester, year of batch (see sample attached)
 - Name of the instructors teaching the course
 - Name of course leader
- Forwarding by HOD bearing his/her signature for approval by Director
- Logo of BVIMR, name of the institution, address
- Warning "strictly for internal use" must be printed on the front title page.
- Table of content bearing
 - Serial no.
 - Contents
 - Page no.
- Copy of latest syllabus of course as specified by university
- Lesson plan bearing
 - Introduction to course
 - Course objectives
 - Learning outcomes
- List of topics/ modules with content
- Evaluation criteria
 - CES evaluation description
 - Recommended text books & reference books
 - Internet resources
 - Swayan courses
- Session plan bearing
 - Session number
 - Topic
 - Readings/case required
 - Pedagogy followed
 - Learning outcome
- Contact details of instructors along with profile
- Main body of course pack having reading material, exercises, case studies, pages for notes
- University question papers (preferably last five years including latest university paper)
- Internal question papers (internal-I-05 papers), (Internal-II-05 papers with latest last year papers)

Note: Include question paper of same subject of old syllabus if required to cover up five years papers.

Declaration by Faculty

I, Tripti Tiwari, Designation Assistant Professor Teaching Business Analytics subject in BCA course VIth sem have incorporated all the necessary pages section/ indicative questions/MCQs mentioned in this check list above.



Tripti Tiwari

BVIMR SNAPSHOT

Established in 1992, Bharati Vidyapeeth (Deemed to be University) Institute of Management and Research (BVIMR), New Delhi focuses on imbibing the said values across various stakeholders through adequate creation, inclusion and dissemination of knowledge in management education. The institute has over the past few years emerged in the lead with a vision of Leadership in professional education through innovation and excellence. This excellence is sustained by consistent value enhancement and initiation of value-added academic processes in institute's academic systems.

Based on the fabulous architecture and layout on the lines of Nalanda Vishwa Vidyalaya, the institute is a scenic marvel of lush green landscape with modern interiors. The Institute which is ISO 9001:2015 certified is under the ambit of Bharati Vidyapeeth University (BVU), Pune as approved by Govt. of India on the recommendation of UGC under Section 3 of UGC Act vide its letter notification No. F. 9 – 16 / 2004 – U3 dated 25th February, 2005.

Strategically located in West Delhi on the main Rohtak Road, BVIMR, New Delhi has splendid layout on sprawling four acres of plot with 'state-of-art' facilities with all class rooms, Library Labs, Auditorium etc., that are fully air-conditioned. The Institute that has an adjacent Metro station "PaschimVihar (East)", connects the entire Delhi and NCR.

We nurture our learners to be job providers rather than job seekers. This is resorted to by fostering the skill and enhancement of knowledge base of our students through various extracurricular, co-curricular and curricular activities by our faculty, who keep themselves abreast by various research and FDPs and attending Seminars/Conferences. The Alumni has a key role here by inception of SAARTHI Mentorship program who update and create professional environment for learners centric academic ambiance and bridging industry-academia gap.

Our faculty make distinctive contribution not only to students but to Academia through publications, seminars, conferences apart from quality education. We also believe in enhancing corporate level interaction including industrial projects, undertaken by our students under continuous guidance of our faculty. These form the core of our efforts which has resulted in being one of the premier institutes of management.

At BVIMR, we are imparting quality education in management at Doctorate, Post Graduate and Under Graduate levels.

PROFILE OF THE COURSE INSTRUCTOR

Ms. TRIPTI TIWARI
Assistant Professor, BVIMR

She is an MBA with specialization in Information Systems along with another degree in Marketing & Finance. Ms. Tiwari has a flair for teaching and research in the area of ICT, Innovation and Technology, Business Analytics and like. Currently with 12.5 years of teaching experience and 6 years Industry experience, she is presently working with Bharati Vidyapeeth (Deemed to be University) Institute of Management & Research, New Delhi.

Ms. Tiwari has been teaching I.T subjects specifically to UG & PG students of Management courses and Computer Application trying to induce the understanding and application of technology in Industry. She heads the Entrepreneurship Development Cell along with many other academic and cultural societies and clubs to provide overall exposure and skill-enhancement among student. She has published various quality research papers in SCIE-Sci, Scopus, WOS and ABDC journals , has edited and authored many national and International books and has 4 patents in her name which has been published and granted.

INDEX

S.No.	CONTENTS	Page No.
1.	Course Outline	3-10
2.	Compiled Notes & Reference Materials	11-157
3.	CES Parameters & helpful contents	158-162
4.	Sample Question Paper Template	163-185
5.	Scribble sheets	186-190



BCA- SEMESTER VI ; ACADEMIC YEAR

Syllabus

Course Number	Course Name	L-T-P- Credits	Year of Introduction
604	Business Analytics	3L-1T-0P=4C	2018-19
Course Objective :			
<ul style="list-style-type: none">• To gain an understanding of how decision makers use business analytics to formulate and solve business problems and to support Information System based decision making.• To become familiar with the processes needed to develop, report, and analyze business data			
Expected Outcome :			
At the end of this course, student should be able to understand <ul style="list-style-type: none">• Identify and prioritize information & data modelling.• Identify and prioritize threats to information assets.• Define an Geographical information system.• Understand various types of Analytics and its significance.• Understand text & web mining• Applications of business analytics			
References (Books, Websites etc) :			
1. Efraim Turban, Ramesh Sharda : Decision Support and Business Intelligence systems : PHI 8 th Edition			
Suggested MOOC : NPTEL, SWYAM			
Course Plan			
Unit	Contents		
1	Business Analytics & Data Visualization: Business Analytics (BA), Overview of Areas where Business Analytics is applied, OLAP, Reports & Queries, Multidimensionality, Advanced Business Analytics, Data Visualization, Geographical Information system, Real time Business Intelligence Automated Decision support, and Competitive Intelligence, BA & Web, Usage benefits & success		
2	Visualization and Data Issues: Organization of Source of Data, Importance of Data Quality, Dealing with Missing or incomplete data, data classification, Introduction to Data Mining, Data mining process, data mining tools XL MINER.		

3	Data, Text & Web Mining : Data Mining concepts & applications, Data Mining Techniques & Tools, Data Mining Project Processes, Text Mining, Web Mining
4	Applications of Business Analytics : Risk - Fraud Detection and Prediction, Recovery Management, Loss Risk Forecasting, Risk Profiling, Portfolio Stress Testing, Market share estimation and Sensitivity Analysis
5	Loyalty Analytics, Customer Life Time Value, Propensity Analytics, Churn Analytics, Customer Analytics Customer Segmentation, Cross- Sell or Up sell Models
6	Recruitment Analytics, Compensation Analytics, Talent Analytics, Training Analytics, Human Resource Retention Analytics, Workforce Analytics Project Work

Course Overview:

In the modern conditions of the dynamic and rapidly changing business environment organizations pursue goals to improve their adaptation capabilities and increase the efficiency of their decision-making process. Due to this organizations implement the Business Intelligence (BI) solutions (reporting, analyzing and forecasting information systems) into their operations. However, the implementation of the BI differs from implementation of a typical IT project due to the fact that business benefits that BI systems provide are intangible and difficult to quantify, companies may face significant problems related to the project planning, cost-benefit analysis, vendor selection, data preparation, employees training and ethical and legal aspects of the implementation.

Many small businesses are reluctant to put BI into practice. Not only because it can be expensive and time consuming to set up, but because they are not clear about the benefits of using it some of which are:

- It's much easier to make informed decisions.
- It's a structured way for growing revenue.
- It increases the competitive advantage over other players in the industry, including larger businesses.
- It improves the efficiency of their business operations.
- It enhances the quality of their customer service.

Business intelligence provides a spectrum of broad range analytical applications, including collaborative BI, mobile BI, open source BI, SaaS BI, operational BI and real-time BI. The technology is not just about gathering intelligence but about making sense of data in a way that can be quickly grasped.

This course will equip the students with theoretical and practical aspects which are now necessary in the corporate world. **The students are expected to review the course readings and the indicated portion of the prescribed text for class discussions prior to attending each session.**

Learning Outcomes:

After undergoing this course, the student will be able to:

- 1: Understand the evolution and basic principle of Business Analytics & Intelligence.
- 2: Recognize when information is needed and how to use it for decision making
- 3: Be able to go through large amounts of data and locate specific pieces of information to use their findings for organizational benefits
- 4: Understand and implement various technologies which help in successful Analytics based decision making
- 5: Comprehend the importance of various types of Analytics with an understanding of how they are utilized in various organizations and global enterprises for various domains.

List of Topics/ Modules:

Topic/ Module	Contents/ Concepts
Module I: Business Analytics & Data Visualization	<ul style="list-style-type: none">• What and why is Business Analytics (BA)• Overview of Areas where Business Analytics is applied• OLAP, Reports & Queries• Concept of Multidimensionality• Advanced Business Analytics & Data Visualization• Geographical Information system• Real time Business Intelligence Automated Decision support, and Competitive Intelligence• BA & Web, Usage benefits & success Intelligence.
Module II: Visualization and Data Issues	<ul style="list-style-type: none">• The distinction between intelligence, information and data• Organization of Source of Data, Importance of Data Quality• Dealing with Missing or incomplete data, data classification• Introduction to Data Mining, Data mining process, data mining tools XL MINER.
Module III: Data, Text & Web Mining	<ul style="list-style-type: none">• Data Mining concepts & applications• Data Mining Techniques & Tools• Data Mining Project Processes• Text Mining, Web Mining
Module IV: Applications of Business Analytics (Finance and related domain)	<ul style="list-style-type: none">• Risk - Fraud Detection and Prediction• Recovery Management• Loss Risk Forecasting• Risk Profiling

	<ul style="list-style-type: none"> • Portfolio Stress Testing • Market share estimation • Sensitivity Analysis
Module V: Applications of Business Analytics (Customer and Market related domain)	<ul style="list-style-type: none"> • Overview of Customer Loyalty Analytics • Customer Life Time Value • Propensity Analytics • Churn Analytics • Customer Analytics Customer Segmentation, Cross-Sell or Up sell Models
Module VI: Applications of Business Analytics (HR and related domain)	<ul style="list-style-type: none"> • Overview of Analytics in HR function and Recruitment Analytics • Overview of Compensation Analytics

Evaluation Criteria:

SN	Type	Description	marks
1	Internal Exam	2 exams	10+10
2	Case study Analysis	CES- 1	5
3	Class Test	CES-2	5
4	Quiz	(using Moodle)-CES3	5
5	Attendance	75% and above	10
	Total Internal Marks	-	40

**

** If a student attempts all 3 CES, best of 2 will be taken for evaluation out of 10 marks whereas if a student misses 1 CES, the remaining two will have a weightage of 3.33 each.

Recommended/ Reference Text Books and Resources:

Text Books	Efraim Turban, Ramesh Sharda : Decision Support and Business Intelligence systems : PHI 8 th Edition
References	B.I. Practices, Technologies and Management, Rajiv Sabherwal, Becena Fernandez
Internet Resource	<ul style="list-style-type: none"> ➤ https://www.solver.com/xlminer-data-mining ➤ https://www.xlminer.com/

Session Plan:

Module I: Business Analytics & Data Visualization			
Session	Topic	Requirements: Readings/ Cases	Learning Outcome
1	Introduction to Business Intelligence and Business Analytics	Text book 1 Class discussion	To develop a basic understanding of the concepts LO1
2	Understand capabilities & Need of Analytics	Text book 1, Addl.Read: Compiled Notes	To identify key elements of BI. LO1
3	Understanding the difference between Business Intelligence and Business Analytics	Text book 1 Class discussion through ppt.	To develop a basic understanding of the concepts LO1
4	Benefits & Importance of Analytics for Business	Text book 1 Class discussion through ppt.	To develop a basic understanding of the concepts LO1
5	Info. System inability leading to BA	Text book 1, Class discussion Through live cases	To develop a basic understanding of the concepts LO1
6	Different types of Information System for effective Decision making	Text book 1, Class discussion Through live cases	To understand the decision making process LO2
7	Application of B.A in organization to effect decisions	Text book 1, Addl.Read Ref. book 3	To develop a basic understanding of the concepts LO1
8	Understanding how information effects Decision making	Text book 1, Addl.Read Ref. book 3	To apply the understood concepts LO1
9	Concept of OLAP and OLTP	Text book 1, Addl.Read: Compiled Notes	To recognize & accordingly apply the concepts LO3
10	Types of OLAP& OLTP	Text book 1, Addl.Read: Compiled Notes	To understand working of the decision making process LO2
11	Reports and Query for Business Decisions	Text book 1, Addl.Read: Compiled Notes	To understand working of the decision making process LO2
12	Analytical Decision Making models & its need	Text book 1, Addl.Read: Compiled Notes	To understand working of the decision making process LO2
13	Multidimensionality & Advanced B.A	Text book 1, Addl.Read: Compiled Notes	To demonstrate the relevance of data LO3

CES 1 : Class Test			
14	Data Visualisation and its significance in Decision making	Text book 1, Addl.Read: Compiled Notes	To understand the use of data by Organization LO3
15	Understanding Data Mining	Text book 1, Addl.Read: Compiled Notes	To understand the use of data by Organization LO3
16	Concept of Data Mining and G.I. S	Text book 1, Addl.Read: Compiled Notes	To understand the use of data by Organization LO3
17	Application of Geographical Information system	Text book 1, Addl.Read: Compiled Notes	To understand the use of data by Organization LO3
18	Real Time B.A and Automated Decision Support	Text book 1, Addl.Read: Compiled Notes	To understand the use of data by Organization LO3
19	Automated Decision Support & Competitive Intelligence	Text book 1,Ch 5	To understand the importance of data for Organization LO3
20	Competitive Intelligence and Web Analytics, Web Analytics and B.A		
Module II: Visualization and Data Issues			
21	Organization of Source of Data & Assessing Data Quality	Text book 1, Addl.Read: Compiled Notes	To demonstrate the need and importance of Information LO4
22	Finding quality Information based on necessary Dimensions	Text book 1, Addl.Read: Compiled Notes	To demonstrate the need and importance of Information LO4
23	Organising Data Sources- Case and Examples	Text book 1, Addl.Read: Compiled Notes	To demonstrate the need and importance of Information LO4
24	Information Asymmetry and its effect on decision making, Dealing with Missing or incomplete data	Text book 1, Addl.Read: Compiled Notes	To understand the importance of Information in decision making LO4
Module III: Data, Text & Web Mining			
25	Overview of Data Mining Process	Text book 1, Addl.Read: Compiled Notes	To demonstrate & apply the various models enhancing importance of Information LO3
26	Data Mining Tools & Techniques	Text book 1,Ch 5	To create Information for analytical decision making LO4

27	Text Mining, Web Mining Concepts	Addl.Read: Compiled Notes	To understand how to handle the critical information LO3
CES 2: Case Study Analysis			
Module IV: Applications of Business Analytics for Finance			
28	Application of B.A in Finance and related domain -an overview	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
29	B.A applications detection and prediction for financial decisions	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
30	Risk and fraud detection and prediction-Real case discussion	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
31	Recovery Management & Business loss and risk forecasting -Real case discussion	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
32	Portfolio Stress Testing & Market share estimation and Sensitivity Analysis -Real case discussion	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
Module V: Applications of Business Analytics for Customer and Market Analytics			
33	Concept of Loyalty Analytics	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
34	Customer Life time value Analytics	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
35	Propensity Analytics, Churn Analytics	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
36	Customer Analytics Customer Segmentation, Cross- Sell or Up sell Models	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
Module VI: Applications of Business Analytics for HR and related Analytics			
37	Overview of Analytics in HR function and Recruitment Analytics	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and is shared LO3
38	Overview of Compensation Analytics	Text book 1, Addl.Read: Compiled Notes	To understand how information flows and its relevance in decision making LO3

39	Overview of Talent Analytics and its application	Text book 1, Addl.Read: Compiled Notes	To learn how customer data can provide better insights LO5
40	Overview of Training Analytics and its application	Text book 1, Addl.Read: Compiled Notes	To understand how to handle the critical information LO3
41	Overview of Human Resource Retention Analytics & Workforce Analytics Project Work	Text book 1, Addl.Read: Compiled Notes	To learn working of technologies LO5
<i>Overall Revision & CES 3: Quiz</i>			

For Internal Circulation at BVIMR Only

COMPILED NOTES & REFERENCE MATERIALS

For Internal Circulation
at BVIMR only

Subject: Business Analytics

Subject Code: 604

Course: BCA Sem-VI

UNIT-I: Business Analytics & Data Visualization

What is Analytics?

Big Data cannot be converted into an asset unless it is analyzed and insights are mined from it. This is where Big Data Analytics comes into the picture. The process of mining useful information (i.e. relevant and useful insights from raw data) from the plethora of data being generated to make smart business decisions, is Big Data Analytics. (This is how the word —information| differs from the word —data|- other pair of words that are used interchangeably.)

Analytics is a process of discovery, interpretation, and communicating meaningful patterns in data. It denotes a persons' skill to gather and use data to generate insights that lead to fact-based decision making. Data-driven analytics provides us with unparalleled opportunities that will help to transform the vast areas concerning business, healthcare, government, etc. The application of data-driven analytics is especially valuable in areas rich with recorded information. Analytics banks on the simultaneous application of statistics, computer programming, and operation research to measure performance. It is observed that analytics most likely favours data visualization while communicating insight.

Analytics also supports the organizations to use the generated business data. It helps the organizations to describe, predict, and enhance their business performance. Business analytics makes extensive use of statistical analysis, including explanatory and predictive modeling, and fact-based management to drive decision making. It is therefore closely related to management science. Analytics may be used as input for human decisions or may drive fully automated decisions. Business intelligence is querying, reporting, online analytical processing (OLAP), and "alerts."

In other words, querying, reporting, OLAP, and alert tools can answer questions such as what happened, how many, how often, where the problem is, and what actions are needed. Business analytics can answer questions like why is this happening, what if these trends continue, what will happen next (that is, predict), what is the best that can happen (that is, optimize)

Examples of Application

Banks, such as Capital One, use data analysis (or analytics, as it is also called in the business setting), to differentiate among customers based on credit risk, usage and other characteristics and then to match customer characteristics with appropriate product offerings. Harrah's, the gaming firm, uses analytics in its customer loyalty programs. E & J Gallo Winery quantitatively analyzes and predicts the appeal of its wines. Between 2002 and 2005, Deere & Company saved more than \$1 billion by employing a new analytical tool to better optimize inventory. A telecom company that pursues efficient call centre usage over customer service may save money.

Types of Analytics

- **Decision analytics:** supports human decisions with visual analytics the user models to reflect reasoning
- **Descriptive analytics:** gains insight from historical data with reporting, scorecards, clustering etc.
- **Predictive analytics:** employs predictive modeling using statistical and machine learning techniques
- **Prescriptive analytics:** recommends decisions using optimization, simulation, etc.

Business Intelligence

Introduction

Business Intelligence (BI) promises a range of technologies for using information within organizations to ensure compliance to strategic and tactical objectives, as well as government laws and regulations. As a research field, BI encompasses data and knowledge management, modeling of processes and policies, data quality, data privacy and security, data integration, data exchange, data cleaning, inconsistency management, information retrieval, data mining, analytics, and decision support.

Dynamic nature of the present business environment requires organizations to quickly respond to external changes (pressures and opportunities) and efficiently adapt their corporate strategies. An efficient decision-making process in business requires a large amount of data, information and knowledge for evaluation and analysis of the possible courses of actions (Turban et al., 2011: 23). Rapid technological development in the database management field during the past decades has allowed businesses to store large amount of historical data which can be used to support the decision-making process by application of the computerized analytical and data processing tools, which are nowadays referred to as business intelligence (BI).

Analytics way before the Computers“ age

The analysis of data led to knowledge discovery for hundreds of years now. So, starting from the data collection project by the Swedish government in 1749 to Florence Nightingale recording, And analyzing mortality data in the 1850s, to British scholar Richard Doll’s tobacco and lung cancer study in the 1950s.

Each of these systems has empowered the field by responding to questions, the answers to which were unknown to us. The Swedes sought answers to the geographical distribution of their population to learn the most efficient way to sustain an appropriate military force. Nightingale wanted to know the role that hygiene and nursing care played in mortality rates. Whereas Doll wished to know if people who smoked had more chances of suffering from lung cancer.

What does Analytics answer?

Analysis of data can reveal correlations and patterns. With the data analytics in the picture, there becomes a less need for the people to rely on hunches and intuitions. Also, the analysis of data can help us answering the following questions:

- Descriptive: What has happened?
- Diagnostic: Why did it happen?
- Predictive: What is likely to happen?
- Prescriptive: Is there anything I can do about it?

Analysis and Analytics – the same concept?

Data Analysis is a broad spectrum that includes Analysis of all kinds, on data sets of all sizes. At a basic level, working with functions and formatting data in Microsoft Excel is an example of Analysis.

Excel was the tool largely used by businesses for a long time. But as the volume of data grew, Excel couldn't be relied upon as a does it all tool. Analysis tools had to be scaled to fit "bigger" data as well. Therefore, new tools had to be developed to deal with Big Data. This led to the birth of Hadoop.

Analysis largely deals with analysing past data and understanding the data. Analytics deals with using these insights to make smart business decisions in the future.

Basic Domains within Analytics

- Behavioral analytics
- Cohort Analysis
- Collections analytics
- Contextual data modeling - supports the human reasoning that occurs after viewing "executive dashboards" or any other visual analytics
- Cyber analytics
- Enterprise Optimization
- Financial services analytics
- Fraud analytics
- Marketing analytics
- Pricing analytics
- Retail sales analytics
- Risk & Credit analytics
- Supply Chain analytics
- Talent analytics
- Telecommunications
- Transportation analytic

Challenges of business analytics

Business analytics has some potential pitfalls that you'll need to overcome.

For starters, you'll find the most success with it when all parties within your company fully support its adoption and execution. It's always going to need buy-in from senior leadership and a clear corporate strategy.

Getting everyone in upper management to sign off on a BA strategy can be difficult, so be sure to present business analytics as supportive of strategies already in place. This should also include clear and measurable goals to help those who are slow to be convinced by the benefits of BA.

In addition to executive ownership, business analytics also require IT involvement, meaning the right technology infrastructure and tools in place to handle the data. Business and IT teams must work together for business analytics to truly succeed. While you're at it, make sure you have the right project management software in place to implement predictive models and adopt an agile approach.

During the initial months of an analytics project, it's important to remain committed to the end result. While the cost of analytics software may be high, and ROI isn't immediate, stay dedicated. The analytical models will develop over-time and predictions will only improve. A business that isn't able to make it through the investment period will likely abandon the entire concept.

Once your analytics are presented, you'll also need end-user buy-in.

End-users need to be involved in adopting business analytics and have a stake in the predictive model that was developed. With that comes top-notch change management, as your organization should be prepared for the changes that these insights will bring to current business and technology operations.

Careers in business analytics

A career in business analytics is a popular choice among those who enjoy working with numbers. To start working towards a career in BA, you'll need a bachelor's degree in business analytics, data science, information management, business intelligence, marketing, statistics, or a related field.

Some of the more popular career paths related to business analytics are:

- **Data Analyst or Data Scientist:** As a data scientist, you would collect, analyze, and organize data in a way that provides the organization with valuable insights that can be utilized by all departments. A data analyst presents this data to upper management using tables, charts, and other types of reports.
- **Business Intelligence Analyst:** A business intelligence analyst is different in the way that they will be gathering and analyzing information to gain an advantage over

competing organizations. They'll present to upper management exactly where their business stands, its strengths and weaknesses, and how they can bring in a larger profit.

- **Big Data Analytics Specialist:** Using the latest developments in technology and data science, big data analytics specialists solve challenges that arise when working within a digital industry. They will often be asked to weigh in on various decisions using insights gained through data and need to be able to backup their conclusions with factual evidence.
- **Management Analyst or Consultant:** The role of a managing analyst consists of working with business operations and making sure they're running smoothly and effectively. You'll work with several other departments to narrow down which business process needs to be improved while also finding a way to enhance efficiency.
- **Marketing Manager:** Those who choose the route of a marketing manager will be required to come up with the marketing strategies of the organization. Whether that's overseeing marketing campaigns, gathering retail analytics, working directly with the sales and marketing teams, or reporting to upper management, it'll likely depend on the type of organization and industry.
- **Operations Research Analyst:** Operation research analysts work to analyze operational data using information technology to run an analysis and develop solutions to improve efficiencies across varying departments.
- **Market Research Analyst:** Those who choose to be a market research analyst will work directly with marketing data. This type of information will help to identify potential customers, evaluate the desirability of the product, and develop specific price ranges to increase revenue over time.

Introduction to Decision Making, Different Styles of Decision and Herbert Simon Model

Decision making

Decision making is the mental processes (cognitive process) resulting in the selection of a course of action among several alternative scenarios.

Working with people is an art and the manager adopts different styles for their own personality, the culture of the organization, and the nature of the people they are working with.

When we look at decision making, the style used will also vary depending upon the nature of the situation and the decision that needs to be made..

The directive style, sometimes referred to "autocratic" style, reflects an individual style where the decision maker relies on their own information, knowledge, experience and judgment. The other three styles of decision making entail varying degrees of involvement of others in gathering information and perspectives, and may include a direct role in making the decision.

Steps of Decision Making Process:

Decision making is a daily activity for any human being. There is no exception about that. When it comes to business organizations, decision making is a habit and a process as well.

Effective and successful decisions make profit to the company and unsuccessful ones make losses. Therefore, corporate decision making process is the most critical process in any organization.

In the decision making process, we choose one course of action from a few possible alternatives.

In the process of decision making, we may use many tools, techniques and perceptions.

In addition, we may make our own private decisions or may prefer a collective decision.

Usually, decision making is hard. Majority of corporate decisions involve some level of dissatisfaction or conflict with another party.

Following are the important steps of the decision making process. Each step may be supported by different tools and techniques.

Step 1: Identification of the purpose of the decision:

In this step, the problem is thoroughly analysed. There are a couple of questions one should ask when it comes to identifying the purpose of the decision.

- What exactly is the problem?
- Why the problem should be solved?
- Who are the affected parties of the problem?
- Does the problem have a deadline or a specific time-line?

Step 2: Information gathering:

A problem of an organization will have many stakeholders. In addition, there can be dozens of factors involved and affected by the problem.

In the process of solving the problem, you will have to gather as much as information related to the factors and stakeholders involved in the problem. For the process of information gathering, tools such as 'Check Sheets' can be effectively used.

Step 3: judging the alternatives:

In this step, the baseline criteria for judging the alternatives should be set up. When it comes to defining the criteria, organizational goals as well as the corporate culture should be taken into consideration.

As an example, profit is one of the main concerns in every decision making process. Companies usually do not make decisions that reduce profits, unless it is an exceptional case. Likewise, baseline principles should be identified related to the problem in hand.

Step 4: Analyse the different choices:

For this step, brainstorming to list down all the ideas is the best option. Before the idea generation step, it is vital to understand the causes of the problem and prioritization of causes. For this, you can make use of Cause-and-Effect diagrams and Pareto Chart tool. Cause-and-Effect diagram helps you to identify all possible causes of the problem and Pareto chart helps you to prioritize and identify the causes with highest effect.

Then, you can move on generating all possible solutions (alternatives) for the problem in hand.

Step 5: Evaluation of alternatives:

Use your judgement principles and decision-making criteria to evaluate each alternative. In this step, experience and effectiveness of the judgement principles come into play. You need to compare each alternative for their positives and negatives.

Step 6: Select the best alternative:

Once you go through from Step 1 to Step 5, this step is easy. In addition, the selection of the best alternative is an informed decision since you have already followed a methodology to derive and select the best alternative.

Step 7: Execute the decision:

Convert your decision into a plan or a sequence of activities. Execute your plan by yourself or with the help of subordinates.

Step 8: Evaluate the results:

Evaluate the outcome of your decision. See whether there is anything you should learn and then correct in future decision making. This is one of the best practices that will improve your decision-making skills.

Conclusion

When it comes to making decisions, one should always weigh the positive and negative business consequences and should favour the positive outcomes.

This avoids the possible losses to the organization and keeps the company running with a sustained growth. Sometimes, avoiding decision making seems easier; especially, when you get into a lot of confrontation after making the tough decision.

But, making the decisions and accepting its consequences is the only way to stay in control of your corporate life and time.

TECHNIQUES OF DECISION MAKING

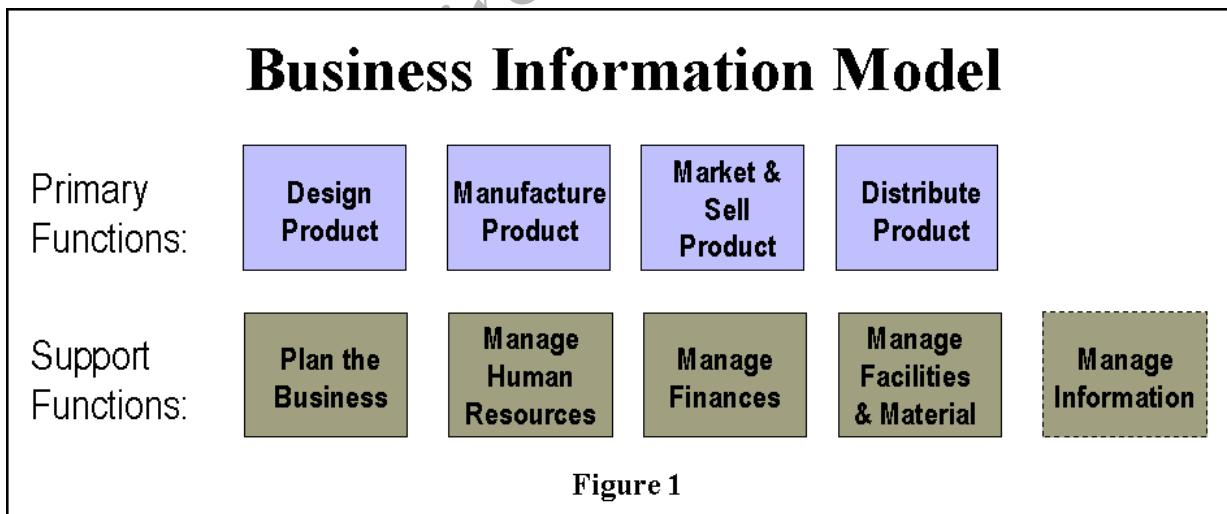
Programmed and Non Programmed Decisions

Let us distinguish two polar types of decisions and call them **programmed decisions** and **non-programmed decisions**, respectively. Having christened them, let add that they are not really distinct types, but a whole continuum, with highly programmed decisions at one end of that continuum and highly unprogrammed decisions at the other end.

Decisions are **programmed** to the extent that they are repetitive and routine, to the extent that a definite procedure has been worked out for handling them so that they don't have to be treated de novo each time they occur. The obvious reason why programmed decisions tend to be repetitive, and vice versa, is that if a particular problem recurs often enough, a routine procedure will usually be worked out for solving it. Numerous examples of programmed decisions in organizations will occur to you: pricing ordinary customers' orders; determining salary payments to employees who have been ill; reordering office supplies.

How Do We Apply Information Systems to Businesses?

- Many companies use their own Business Information Model (BIM) to define the major business functions of their companies.
- In order to define a model, they divide their business functions into 2 main groups.

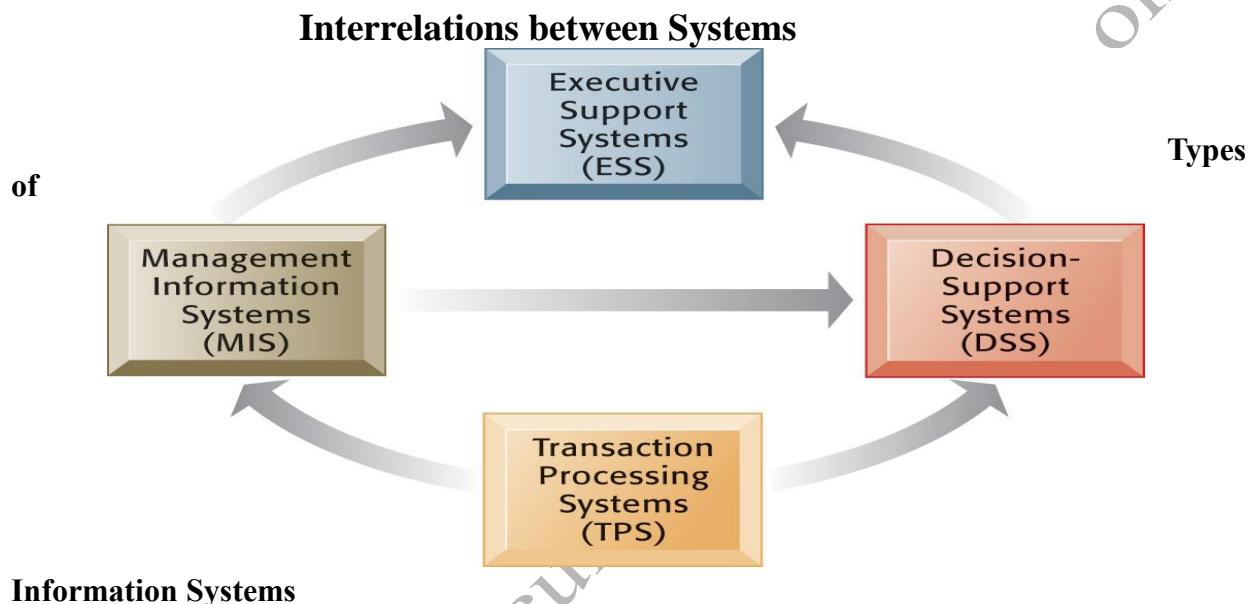


Primary functions are needed to develop and deliver products or services.

- Support functions are needed to perform in order to support primary functions.

Relationship Between Systems

- TPS: Major source of data for other systems
- ESS: Primarily a recipient of data from lower-level systems
- Other systems may exchange data as well
- Exchange of data between functional areas
 - E.g. Sales order transmitted to manufacturing system
- In most organizations, systems are loosely integrated



A management information system (MIS) is a computer-based system that provides the information necessary to manage an organization effectively. An MIS should be designed to enhance communication among employees, provide an objective system for recording information and support the organization's strategic goals and direction. There are four types of MIS that will be introduced in ascending order of sophistication.

Transaction Processing Systems

- These systems are designed to handle a large volume of routine, recurring transactions. They were first introduced in the 1960s with the advent of mainframe computers. Transaction processing systems are used widely today. Banks use them to record deposits and payments into accounts. Supermarkets use them to record sales and track inventory. Managers often use these systems to deal with tasks such as payroll, customer billing and payments to suppliers.

Operations Information Systems

- These systems were introduced after transaction processing systems. An operations information system gathers comprehensive data, organizes it and summarizes it in a form that is useful for managers. These types of systems access data from a transaction processing system and organize it into a usable form. Managers use operations information systems to obtain sales, inventory, accounting and other performance-related information.

Decision Support Systems (DSS)

A DSS is an interactive computer system that can be used by managers without help from computer specialists. A DSS provides managers with the necessary information to make intelligent decisions. A DSS has three fundamental components:

1. Database management system (DBMS): Stores large amounts of data relevant to problems the DSS has been designed to tackle.
2. Model-based management system (MBMS): Transforms data from the DBMS into information that is useful in decision making.
3. Dialog generation and management system (DGMS): Provides a user-friendly interface between the system and the managers who do not have extensive computer training.

Expert Systems and Artificial Intelligence

These systems use human knowledge captured in a computer to solve problems that ordinarily need human expertise. Mimicking human expertise and intelligence requires that the computer (1) recognize, formulate and solve a problem; (2) explain solutions and (3) learn from experience. These systems explain the logic of their advice to the user; hence, in addition to solving problems they can also serve as a teacher. They use flexible thinking processes and can accommodate new knowledge.

Types of Decision

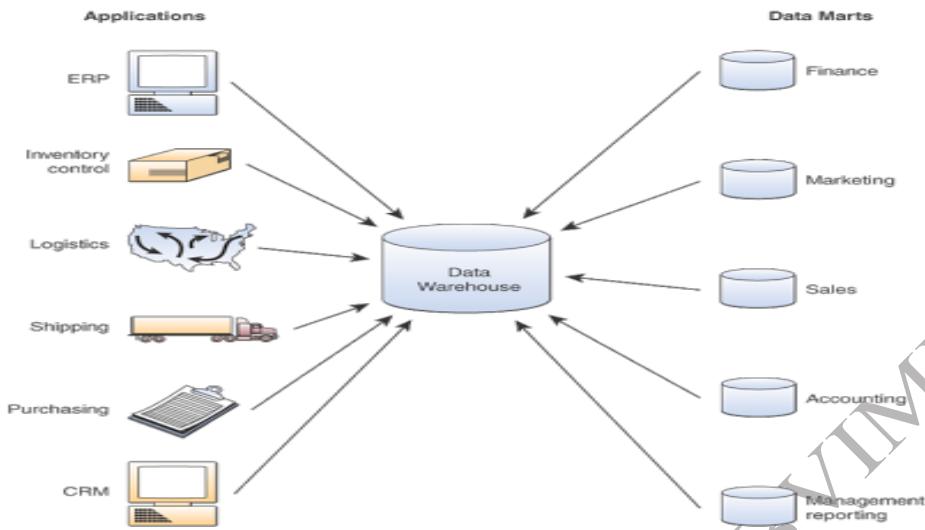
A manager can make two kinds of decision:

- **Structured** – which are repetitive and need a definite routine and procedure to deal with them, e.g. stock is below 15 %, so an order need to be place with a supplier.
- **Unstructured** – require knowledge, insight, and evaluation. They may well crop up without warning, and the right decision can be critical. Eg: Whether the organization should think about diversifying now!!

Data Warehouse

A data warehouse stores data from current and previous years that has been extracted from the various operational and management databases of an organization. It becomes a central source of data, which has been screened, edited, standardized, and integrated so it can be used by

managers and other end user professionals throughout an organization. Data warehouses may be subdivided into data marts, which hold specific subsets of data from the warehouse.



Data Mining

A major use of data warehouse databases is data mining. In data mining, the data in a data warehouse are processed to identify key factors and trends in historical patterns of business activity that can be used to help managers make decisions about strategic changes in business operations to gain competitive advantages in the marketplace.

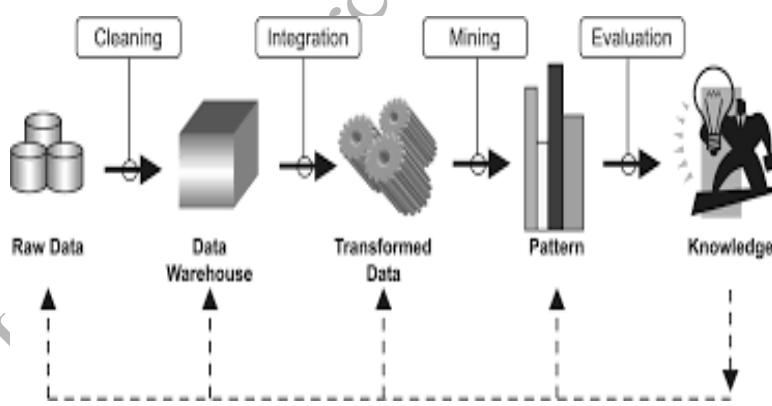
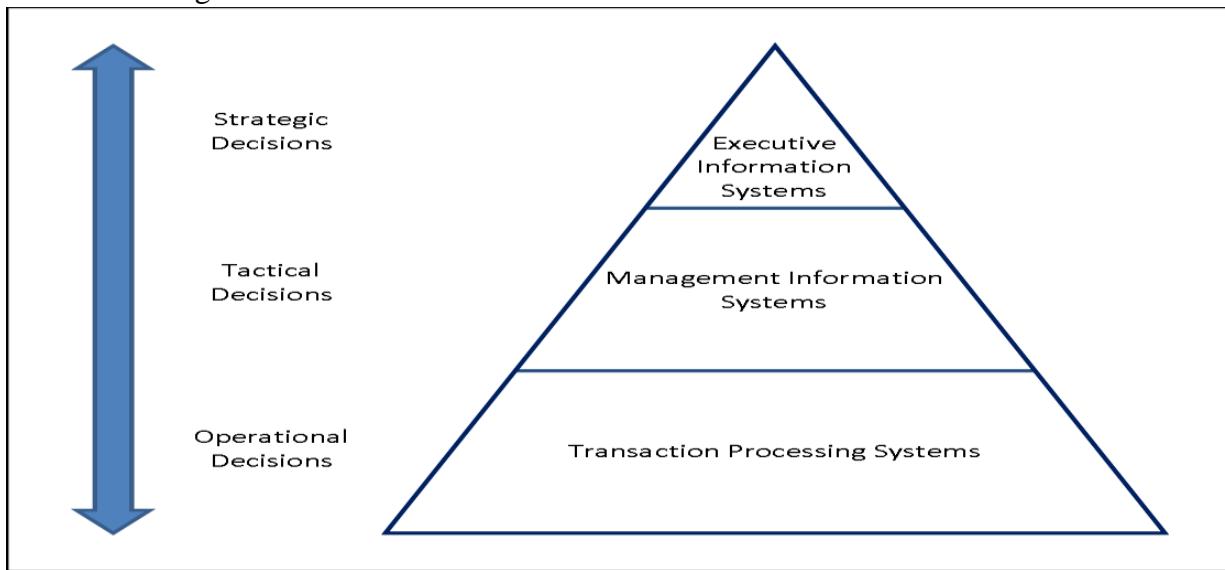


Figure: Data Mining Discovers Patterns in Data

Information Needs of Managers at Different Levels

As most organizations are hierarchical, the way in which the different classes of information systems are categorized tends to follow the hierarchy. This is often described as "the pyramid model" because the way in which the systems are arranged mirrors the nature of the tasks found at various different levels in the organization.

For example, this is a three level pyramid model based on the type of decisions taken at different levels in the organization.



SENIOR MANAGEMENT

Also known as **Strategic Level** – long term decisions that will affect the future of the organisation, e.g. whether to open a new store, or take over a rival concern. -information for strategically positioning the organization -competitive analysis and performance evaluation, -strategic planning and policy, -external factors that influence the direction

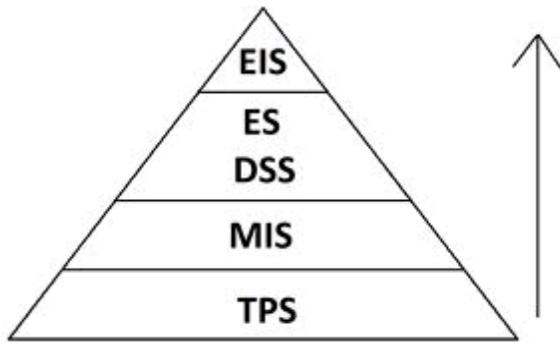
MID LEVEL MANAGEMENT

Also known as **Tactical Level** – decisions that have a short to medium term effect, e.g. introducing a new product to a particular retail outlet Organizational and operational functions -information for coordination of work units -information for delivery programmes -evaluation of resources usage -budget control, problem solving -operational planning

LOWER LEVEL MANAGEMENT

Also known as **Operational Level** – day-to-day decisions such as ordering in more stock Programme management within units -information for implementing programmes -information for managing programmes -management of resources usage -project scheduling -problem solving -operational planning

TYPES OF INFORMATION SYSTEM



Description

Executive Support Systems

An Executive Support System ("ESS") is designed to help senior management make strategic decisions. It gathers, analyses and summarises the key internal and external information used in the business.

A good way to think about an ESS is to imagine the senior management team in an aircraft cockpit - with the instrument panel showing them the status of all the key business activities. ESS typically involve lots of data analysis and modelling tools such as "what-if" analysis to help strategic decision-making.

Management Information Systems

A management information system ("MIS") is mainly concerned with internal sources of information. MIS usually take data from the transaction processing systems (see below) and summarise it into a series of management reports.

MIS reports tend to be used by middle management and operational supervisors.

Decision-Support Systems

Decision-support systems ("DSS") are specifically designed to help management make decisions in situations where there is uncertainty about the possible outcomes of those decisions. DSS comprise tools and techniques to help gather relevant information and analyse the options and alternatives. DSS often involves use of complex spreadsheet and databases to create "what-if" models.

Knowledge Management Systems

Knowledge Management Systems ("KMS") exist to help businesses create and share information. These are typically used in a business where employees create new knowledge and expertise - which can then be shared by other people in the organisation to create further commercial opportunities. Good examples include firms of lawyers, accountants and management consultants.

KMS are built around systems which allow efficient categorisation and distribution of knowledge. For example, the knowledge itself might be contained in word processing documents, spreadsheets, PowerPoint presentations, internet pages or whatever. To share the knowledge, a KMS would use group collaboration systems such as an intranet.

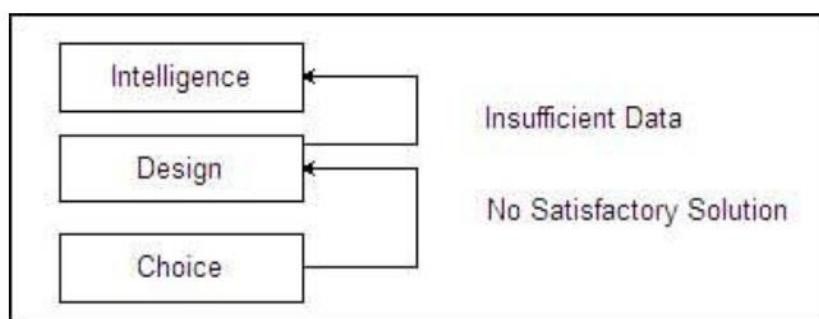
HERBERT SIMON'S MODEL

Herbert Simon made key contributions to enhance our understanding of the decision-making process. In fact, he pioneered the field of decision support systems. According to (Simon 1960) and his later work with (Newell 1972), decision-making is a process with distinct stages. He suggested for the first time the decision-making model of human beings. His model of decision-making has three stages:

1. **Intelligence** which deals with the problem identification and the data collection on the problem.
2. **Design** which deals with the generation of alternative solutions to the problem at hand.
3. **Choice** which is selecting the 'best' solution from amongst the alternative solutions using some criterion.

The figure given below depicts Simon's decision-making model clearly.

The figure given below depicts Simon's decision-making model clearly.



Intelligence Phase:

This is the first step towards the decision-making process. In this step the decision-maker identifies/detects the problem or opportunity. A problem in the managerial context is detecting anything that is not according to the plan, rule or standard. An example of problem is the detection of sudden very high attrition for the present month by a HR manager among workers. Opportunity seeking on the other hand is the identification of a promising circumstance that might lead to better results. An example of identification of opportunity is-a marketing manager gets to know that two of his competitors will shut down operations (demand being constant) for some reason in the next three months, this means that he will be able to sell more in the market.

Thus, we see that either in the case of a problem or for the purpose of opportunity seeking the decision-making process is initiated and the first stage is the clear understanding of the stimulus that triggers this process. So if a problem/opportunity triggers this process then the first stage deals with the complete understanding of the problem/opportunity. Intelligence phase of decision-making process involves:

- a. **Problem Searching**
- b. **Problem Formulation**

Problem Searching: For searching the problem, the reality or actual is compared to some standards. Differences are measured & the differences are evaluated to determine whether there is any problem or not.

Design Phase:

Design is the process of designing solution outlines for the problem. Alternative solutions are designed to solve the same problem. Each alternative solution is evaluated after gathering data about the solution. The evaluation is done on the basic of criteria to identify the positive and negative aspects of each solution. Quantitative tools and models are used to arrive at these solutions. At this stage the solutions are only outlines of actual solutions and are meant for analysis of their suitability alone. A lot of creativity and innovation is required to design solutions.

Choice Phase:

It is the stage in which the possible solutions are compared against one another to find out the most suitable solution. The 'best' solution may be identified using quantitative tools like decision tree analysis or qualitative tools like the six thinking hats technique, force field analysis, etc.

This is not as easy as it sounds because each solution presents a scenario and the problem itself may have multiple objectives making the choice process a very difficult one. Also uncertainty about the outcomes and scenarios make the choice of a single solution difficult.

Extensions to Simon's Model

Implementation

The decision that is ultimately carried out.

Review

In this phase, decision implemented is evaluated. Was the course of action taken a good choice?

DECISION ANALYSIS BY ANALYTICAL MODELLING

Or (Models of Decision Making)

Decisions needs to be analyzed for conditions and assumptions considered in the decision model. The process is executed through analytical modeling of problem and solution. The model is analyzed in four ways:

1. **What if analysis** – Decisions are made using a model of the problem for developing various solution alternatives & testing them for best choice. The model is built with some variables and relationship between variables .in reality, the considered values of variables and relationships may not hold good and therefore solution needs to be tested for an outcome, if considered values of variables or relationship change.

This method of analysis is called 'what if analysis'. **In other words**, observing how changes to selected variables affect other variables.

Example: What if we cut advertising by 10%? Will anything happen to sales?

2. **Sensitivity analysis** – It is a special case of what if analysis in which only one variable is changed and rest are kept unchanged. It helps to understand the significance of variable in decision making and improves the quality of decision making. **In other words**, observing how repeated changes to a single variable affect other variable.

Example: Lets cut advertising by Rs.100 repeatedly so we can see its relationship to sales.

3. Goal seeking analysis – in this, the problem is analyzed in exactly reverse way as that of what if analysis or sensitivity analysis. In goal seeking analysis, the goal is fixed and the variables and values are analyzed, which would help to seek that goal. The work is done backward from the goal. **In other words**, making repeated changes to selected variables until a chosen variable reaches a target value.

Example: Let's try increasing advertising until sales reach 1 million mark.

4. Optimisation analysis - in this, the goal is not fixed but the decision maker tries to achieve a goal of an optimum value arrived at after satisfying all constraints operating in the problem. This method is a more complex extension of Goal-seeking analysis. The decision maker can use this analysis to work on constraints and resources and find ways to improve upon solution to seek highest goal.

In other words, finding an optimum value for selected variables, given certain constraints.

Example: What is the best amount of advertising to have, keeping in mind our budget and choice of media.

Example: What is the best amount of advertising to have, keeping in mind our budget and choice of media.

OLAP vs. OLTP system

- **Online Analytical Processing (OLAP).** The most popular technology in data analysis is OLAP. OLAP servers organize data into multidimensional hierarchies, called cubes, for high-speed data analysis. Data mining algorithms scan databases to uncover relationships or patterns. OLAP and data mining are complementary, with OLAP providing top-down data analysis and data mining offering bottom-up discovery.

OLAP tools allow users to drill down through multiple dimensions to isolate specific data items. For example, a hypercube (the multidimensional data structure) may contain sales information categorized by product, region, salesperson, retail outlet, and time period, in both units and dollars. Using an OLAP tool, a user need only click on a dimension to see a breakdown of dollar sales by region; an analysis of units by product, salesperson, and region; or to examine a particular salesperson's performance over time.

Information can be presented in tabular or graphical format and manipulated extensively. Since the information is derived from summarized data, it is not as flexible as information obtained from an ad hoc query; most tools offer a way to drill down to the underlying raw data. For

example, Power Play provides the automatic launch of its sister product, Impromptu, to query the database for the records in question.

Although each OLAP product handles data structures and manipulation in its own way, an OLAP API, developed by a group of vendors who form the OLAP Council, standardizes many important functions and allows IT to offer the appropriate tool to each of its user groups. The MD-API specifies how an OLAP server and client connect, and it defines metadata, data fetch functions, and methods for handling status messages. It also standardizes filter, sort, and cube functions; compliant clients are able to communicate with any vendor's compliant server.

OLAP Variants: MOLAP, ROLAP, and HOLAP. OLAP is divided into multidimensional OLAP (MOLAP), relational OLAP (ROLAP), and hybrid OLAP (HOLAP).

ROLAP can be applied both as a powerful DSS product, as well as to aggregate and pre-stage multi-dimensional data for MOLAP environments. ROLAP products optimize data for multi-dimensional analysis using standard relational structures. The advantage of the MOLAP paradigm is that it can natively incorporate algebraic expressions to handle complex, matrix-based analysis. ROLAP, on the other hand, excels at manipulating large data sets and data acquisition, but is limited to SQL-based functions. Since all organizations will require both complex analysis and analysis of large data sets, it could be necessary to develop an architecture and set of user guidelines that will enable implementation of both ROLAP and MOLAP where each is appropriate.

HOLAP is the newest step in the ongoing evolution of OLAP. HOLAP combines the benefits of both ROLAP and MOLAP by storing only the most often used data in multidimensional cube format and processing the rest of the relational data in the standard on-the-fly method. This provides good performance in browsing aggregate data, but slower performance in "drilling down" to further detail.

BI is built as an Online Analytical Processing system (OLAP), to provide robust analytical capabilities, such as high-speed access to reports, dashboard management and the development of balanced scorecards. BI also comes with advanced analytical features that allow you to view data from different sources on one page, and in the format or perspective you need.

ERP, on the other hand, is an Online Transaction Processing system (OLTP), used to record transactions as and when they take place. The data architecture of ERP software is designed to provide high-speed transaction recording, while keeping disk space utilization at a minimum.

Agility vs. efficiency

Over the past years, there has been a shift of focus in BI - organizations are moving from historical reporting to forecasting and forward planning. Through these future-centric capabilities, BI can make organizations become more agile, allowing them to make strategic-level decisions that take advantage of future conditions.

ERP software, on the other hand, is built to deliver efficiencies to an organization. These efficiencies come in many forms: better interdepartmental communication, IT cost savings and

business process efficiencies. Both Gartner and CIO.com believe that proper ERP implementation can improve an organization's overall performance.

What does this mean for you?

Before you begin choosing tools, first determine your organisation's objectives. Once you know what you are trying to achieve, you can identify the right approach to help you achieve it.

If you have an in-depth understanding of your operational performance, then look at BI to obtain strategic level insights into your performance. While if you need a better understanding of your operational performance and need to make operational improvements, ERP is the tool you need.

Ultimately, both tools are geared towards business improvement and can deliver significant results.

Query and Report Writing Technologies

Business Intelligence Software

BI software helps organizations organize and analyze data to make better decisions. This could include internal data from company departments as well as from external sources, such as marketing data services, social media channels or even macroeconomic information.

The BI market is growing rapidly because of the proliferation of data to analyze. Over the past few decades, companies that have deployed Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) and other applications are now sitting on a mountain of data that can be analyzed. In addition, the growth of the Web has increased the demand for tools that can analyze large data sets.

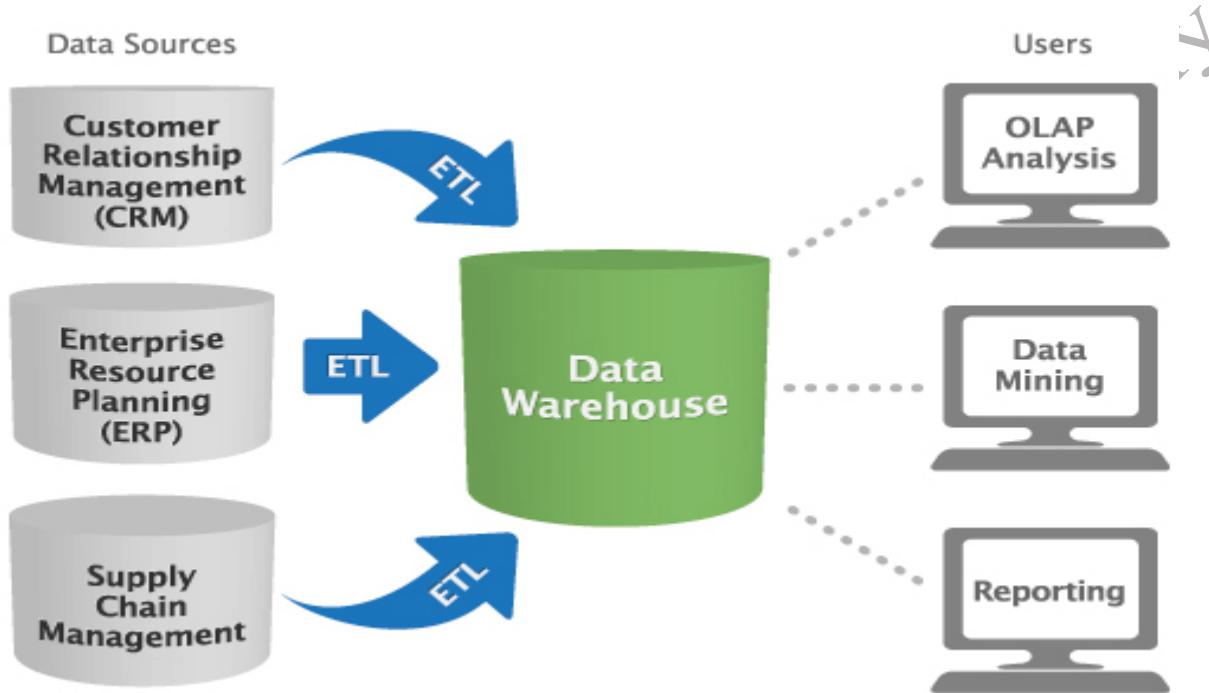
One of the biggest trends in the BI market is the shift in software architecture and design to more user-friendly applications. These applications are now being used by business users—not just IT staff—to analyze particular sets of departmental data, including marketing, procurement, retail and Web data.

Common Features of Business Intelligence Software

BI software can be divided into three broad application categories: data management tools, data discovery applications and reporting tools (including dashboards and visualization software). In the next section, we'll explain how these applications can help your organization's decision-making process become more data-driven.

What BI tools you need depends on how your data is currently managed and how you would like to analyze it. For example, if it is currently scattered across disparate transactional databases, you might need to build a data warehouse to centralize it and invest in data management tools that offer Extract, Transform and Load (ETL) functionality to move and re-structure it.

Once data is given a common structure and format, you can invest in data discovery solutions such as Online Analytical Processing (OLAP), data mining and semantic or text mining applications, with the capability to create custom, ad hoc reports. And because information is stored within the warehouse, users can quickly pull reports without impacting the performance of the organization's software applications, such as CRM, ERP and supply chain management solutions.



Business Query and Reporting

Business query and reporting tools are often referred to as "ad hoc query tools." This terminology is a little misleading, as in fact the queries are not always ad hoc (as in spontaneously crafted) but rather are often fixed reports. The difference is that a business user, usually a power user, may have built the report, rather than an information technology (IT) person. The business environment changes at a rapid pace, and unable to wait weeks or months for IT to develop a new report, business users often demand the ability to create queries and reports themselves. Business query and reporting tools allow for this and are most often used for decision-making and management purposes. The business query and reporting tool is a key module to provide users with self-service information access.

In some cases, a report is truly ad hoc; it's a one-off business question that will never be posed again. Ad hoc queries may be exploratory in nature as users try to find the root cause of a problem, test a theory, or consider changing a business model. Table 3-1 lists some sample fixed

reports that may lead to an ad hoc query. As users explore the data, what started as an ad hoc query or one-time question may later become a fixed report. It's important to recognize the iterative nature of business intelligence and to ensure you have flexible business intelligence tools.

Getting to the data is just one capability of business query tools; the other aspect is presenting and formatting the data in a meaningful way, loosely referred to as *reporting*. The terms "query" and "reporting" are sometimes used interchangeably because a business query and reporting tool will have both capabilities -- getting to the data and formatting it to create a report.

Fixed Report	Purpose	Related Ad Hoc Query
Inventory by Product	To determine if an order can be fulfilled today by the primary warehouse	If I'm short at my main warehouse, can I supply the product from elsewhere?
Top 10 Customers By Quarter and Product	To understand which customers generate the most revenue	Who fell off this quarter's list? Are there certain products we can cross-sell?
Raw Material Receipts and Delivery Times	To determine how long it takes to acquire raw materials and which supplier can fulfill purchase orders fastest	Are there other suppliers who can respond faster?
Patients Per Hour	To understand busy periods and wait times	Do staffing levels correspond to busy times?

Table 3-1 **Sample Fixed and Ad Hoc Reports**

Business query and reporting tools vary widely in their formatting capabilities. The most basic of formatting capabilities allow for changing the font of column headings and making them bold and centered. Conditional formatting will, for example, display numeric values red when negative or below target and green when positive or above target. Simple report styles include displaying information in a cross-tab report, a chart, or a master-detail report with groupings and subtotals. Tools may provide a set of templates to create nicely formatted reports that use a consistent corporate look and feel. More complex formatting capabilities include the ability to present multiple charts on a page, perhaps coming from different data sources.

Examples of business query tools include BusinessObjects Web Intelligence, Cognos 8 Query Studio, and SAS Web Report Studio.

A Business View of the Data

Business query tools allow business users to access a data source via business terms without having to write any SQL. The data source could be a data warehouse as described in Chapter 2, or it might be direct access to an operational system. A key feature of a business query tool is that it has a business view or metadata layer that hides the complexity of the physical database structure from the business user by:

- Using business terminology rather than physical field names. For example, a user may select a dimension such as Customer Name rather than a cryptic field such as CUST.L33_NAME (the physical table and field name in the Relational Database Management System [RDBMS]).
- Automatically connecting related tables via joins.
- Providing metrics that may calculate and aggregate facts such as revenue, number of customers, number of orders, number of incidents, and average selling price.

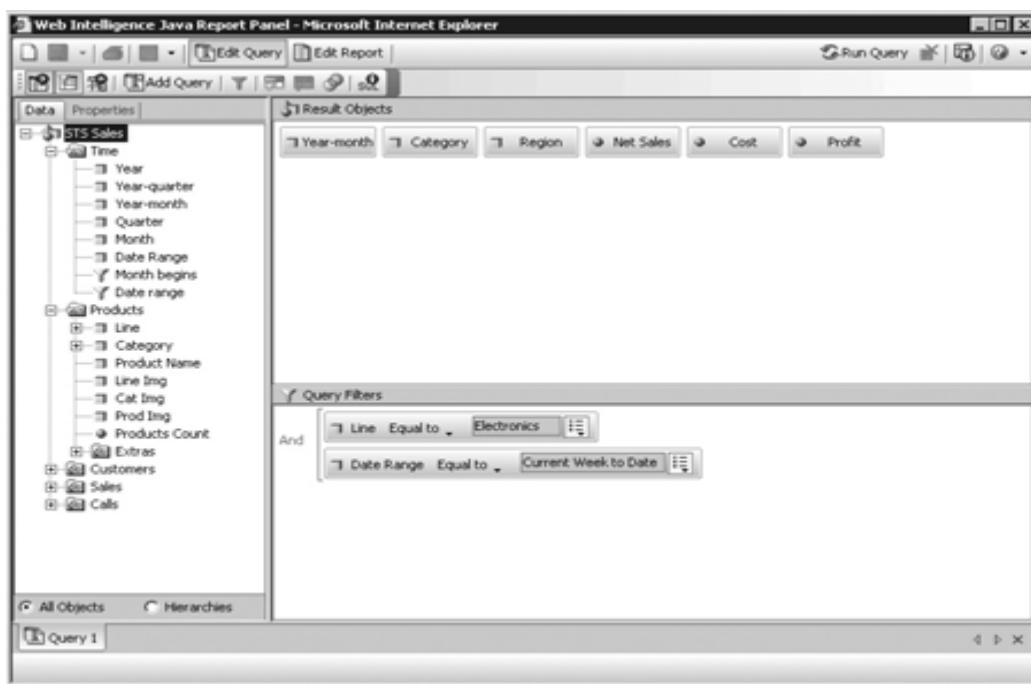


Figure 3.1 shows an example of building a query with the Business Objects universe, one of the first products to introduce the concept of a business view.

This business view is the most important piece of your BI front-end tools and one in which the business and IT must work together to model. For integrated BI platforms, the business view is common to all the BI tool modules: business query, reporting, analysis, and dashboards. When the business view looks too much like the data warehouse or source system with confusing table and field names, business users are overwhelmed and can too easily build incorrect queries. Poor business view design also forces users to put too much logic and too many calculations inside individual reports and dashboards. For these reasons, in some organizations, the power users within a business unit, function, or department, are responsible for building the business view or metadata layer; in others, it is the central BI group or data warehouse team that will build and maintain the business view.

Production Reporting

Whereas business query and reporting tools allow for basic report formatting, production reporting tools have much more sophisticated formatting and design capabilities. Some people

may refer to this category of tools as pixel perfect, operational, or enterprise reporting. Again, the terminology can be misleading as some business query and reporting tools can create pixel perfect reports, be embedded in operational systems, and are used across an enterprise. For lack of a better term, I will refer to this module as "production" reporting. Examples of production reporting tools include Actuate e-Report, Business Objects Crystal Reports, Microsoft Reporting Services, Oracle Publisher (which supersedes Oracle Reports), and Information Builders Web Focus.

A production reporting tool may access a transaction system directly to create a document such as an invoice, a bank statement, a check, or a list of open orders. When the reporting is not against the transaction system, it may be against an operational data store or detailed data within a data warehouse. IT usually develops these reports for the following reasons:

- The data source is an operational system in which you can't take the risk that "untrained" users may launch resource intensive and runaway queries with a business query tool.
- Reports are often accessed through and embedded within the transaction system.
- The information requirements are common to all users and departments and are static, such as for regulatory reports.

Because professional IT developers are often the users of production reporting tools, IT may also use these tools to develop management style reports, particularly when a company does not have a business query tool.

What are reporting tools?

Reporting tools are widely used to support decision making and to measure organizational and team performance. Companies use them for financial consolidation, for evaluation of strategies and policies and often just for plain reporting. Today most of these tools are integrated with Business Intelligence tools.

What are the advantages of these tools for your organization?

Reporting tools allow companies to create attractive reports easily. In tabular or graphical format. With data from Excel, a data warehouse or the organization's ERP system. With the reports containing the right information people are able to manage and improve the business processes more easily.

Which reporting tools are available in the market?

There are many Business Intelligence Reporting tools in the market available. This list of reporting tools shows you a complete overview of all the major tools currently available and thoroughly studied by our industry analysts. To name a few: Cognos Query and reporting, Oracle Enterprise Reporting, Crystal Reports, Microsoft Reporting Services (SSRS), etc.

Is there a difference between reporting tools and BI tools?

Almost every company, large or small, needs a reporting tool or they have already one in place. But, what is the difference between reporting software and Business Intelligence tools? They are not the same. In general reporting tools tell us where we have been but are not very good at analyzing data (at real time) and telling us where we are going (predictive). This is very well shown at the figure below.

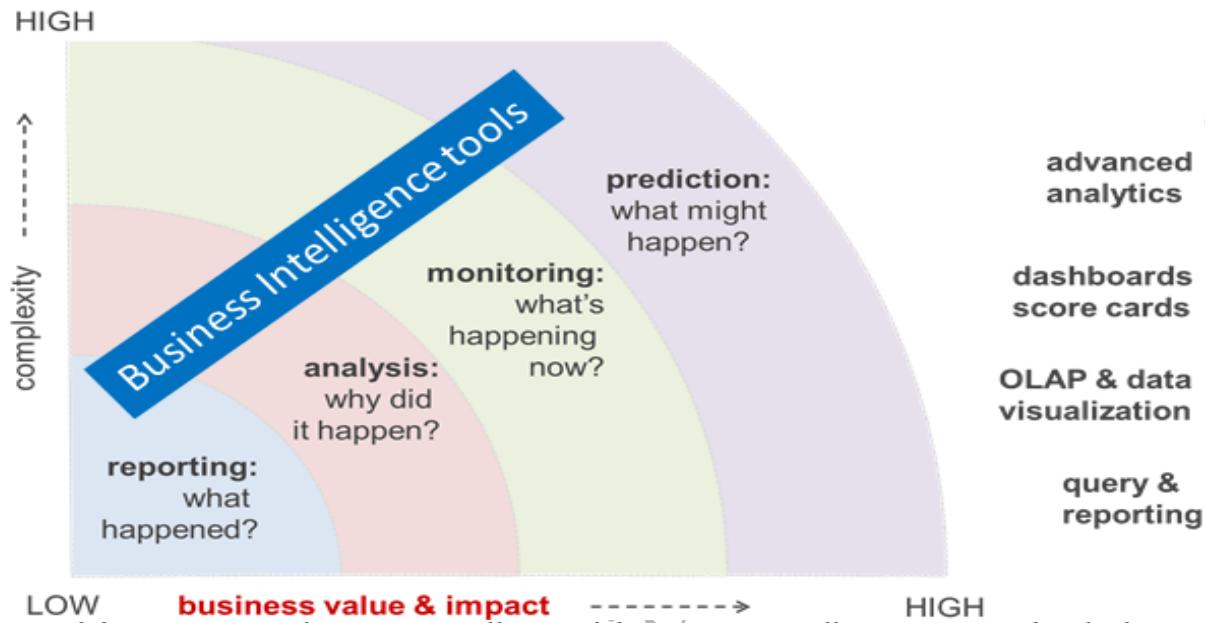


Figure 3.2: reporting tools cover a small part of the Business Intelligence arena, they lack important functionality for Data Discovery and Predictive Analytics.

The figure displays the relationship between the complexity of the different solutions, the business impact and the type of functionality you need. If you need all types of functionality (reporting, OLAP, dashboards, advanced analytics) consider buying a complete Business Intelligence platform.

Are they intelligent?

Almost all Business Intelligence tools contain a reporting tool but most reporting tools are not very smart. That is because they have very less functionality for analyzing data: OLAP or Data Discovery, and they are often not able to process very large volumes of data. Our suggestion is that if you think you need a reporting tool look further and consider a Business Intelligence tool.

A list of some widely used reporting tools

- BusinessObjects BI: Reporter; WebIntelligence; Crystal Reports, v4.1/SP5, SAP
- IBM Cognos Series 10: Cognos Query and Reporting, v10.2.2, IBM
- JasperSoft (open source): iReport Designer, v6, JasperSoft
- Microsoft BI tools: Reporting & Analysis Services; Power BI; SharePoint, 2014, Microsoft
- MicroStrategy: MicroStrategy Report Services, v9.4.1, Microstrategy
- Oracle Enterprise BI Server: Oracle Enterprise Reporting, v11.1.1.6, Oracle

- Oracle Hyperion System: Hyperion Interactive Reporting, v9, Oracle
- Pentaho BI suite (open source): Pentaho Report Designer & Engine, v5.2, Pentaho
- Qlik: Qlik Business Discovery (no pixel-perfect reporting), v11.20, QlikTech
- SAP NetWeaver BI: SAP Crystal Reporting, v7.3, SAP
- SAS Enterprise BI Server: Query & Analysis, v9.4, SAS Institute
- Tableau Software: Does not contain a reporting tool, v8.3, Tableau Software
- WebFocus: Reports & dashboards; Query and Analysis, v8.01, Information Builders

Above is the list of some examples widely used by organisations today. Many operational systems contain a reporting tool, they are embedded. Users are often not aware that they are even there, they just click on a ‘list’ and look into the results. Embedded reporting tools allow users to create reports based on data only of that specific operational system.

The real business value depends on how many data elements can be combined in one report. The more tables that can be joined the bigger the business value might be. This works exponentially if you select the right data elements and the data quality is good. The truth is that many embedded reporting tools have very less business value because the standard reports contain data from just 1 or 2 tables and they often don’t have data cleansing tools on board. However there are a few exceptions.

MULTIDIMENSIONAL BASICS

Information processing is the backbone of a business organization because it is where company-oriented decisions are based. For that reason, good information processing came to be the vital factor for the survival of any business organization. However, good information comes in various states. The need to rely and analyze not only the past and current but also the projected future data is very essential in arriving at a good decision. Just as evaluation of operational performance is a key to maintaining a smoothly running business, so is the ability to make predictions on the possible movements and trends in the industry which normally involves formulation of assumptions, backed by use of derived descriptions, application of certain algorithms and using other information apart from those provided by the actual transactional results. From here, the limiting factors of operational software surfaced, because operational software is unable to provide anything other than historical data. This triggered the birth of several analysis applications in support of analysis-oriented processing.

Using Business Analytics Applications

How then can businesses take advantage of the benefits derived from utilization of these technologies? The system put in place by most companies allows a user to pull, extract, or ‘query’ data based on particular information requirements in support of some ad-hoc analysis directed toward a specific objective. Not only do business organizations put these systems into use for their regular operating activities, but they tend to heavily rely on the information provided by these applications for reporting and analysis purposes.

Typically, SQL databases and Spreadsheet technologies are being employed by most users in analysis and reporting activities. However, technology developers and users have found some flaws and encountered certain difficulties or limitations in using either of the two applications as they do not fully support the analytical requirements of end-users.

For example, the use of spreadsheets is most convenient for the majority of end users, as most people are familiar and have been exposed to this type of application. But although users can move comfortably in a spreadsheet environment, performing an analysis of data in this format can prove to be a very challenging task. Often, users need to refer to multiple numbers of worksheets, particularly when dealing with large volumes of data. Apart from that, performing a complex calculation can also be a major problem, as spreadsheets have only limited calculation functions and features. Spreadsheets are able to handle cell-based formulas, but a user is still required to redefine the formula for the same intersections of data when working on multiple worksheets. Also, spreadsheets also do not fully support hierarchies and multiple hierarchies. Additionally, spreadsheets remain inflexible and do not present an easy and quick method of restructuring views. In instances where a user wants to generate a different view of the data, he will be compelled to redo everything, including redefining formulas and necessary look up tables.

On the other hand, a SQL Database application uses a structured query language to pull specific data from a database source. Unfortunately, many users find it difficult to handle this type of application. As most end users are non-technical people, query language is unnatural for them as it is complex and technically sophisticated. To address this problem, accessible query templates and easy-to-use query tools have been provided. In spite of this, many are still indisposed to use this application because there are still situations where the predefined query templates and tools are not in sync with a user's specific needs. Oftentimes, when dealing with this type of application, users are inclined to be highly dependent on technical support, which suppresses the ability of information to be readily available in a timely manner.

The Concept of Multidimensionality

The problems with analyzing and reporting data from operating systems through SQL highlight the importance of the concept of multidimensionality. In very few words, multidimensionality is the ability to manipulate the view of data in various ways without disrupting the integrity of the data. It is extremely important that users have the ability to customize and change their views of data to suit their specific needs and enable users to perform any type of analysis within the database.

Data or information handled by a small organization varies not only in volume but also in complexity as compared to a medium- to large-sized organization. Large companies would most likely, for instance, have a wider array of products, broader market and consumers, many outlets

in different locations, several suppliers, a larger employee count and large companies handle more denominations and measures. Also since there are different users of information, whom may come from different levels within the organizational structure, they can differ not only in the needed views of data but also have varying reporting requirements and need different reporting schemas.

When we talk about a multidimensional database, the foremost consideration would be the dimension, which happens to be the basic structure of a multidimensional application. Each dimension has a structure, a set of members for each category, and a child/parent relationship between members. Fields of data may be organized into logical categories such as Time, Geography, Customers, Products and Measures, to name a few examples of dimensions. Some dimensions are arranged following a certain hierarchy structure. Thus, in a geographic dimension, local branches may be consolidated into regions, areas and countries. In such case where a hierarchical dimension is present, we will regard the entire hierarchy as one single dimension.

Normally, a user may want to pull only a particular set of data. For example, a sales manager would want to infer how well their sales operations are doing, and he wants to break the information down by products, by location, by customer and by season. Every time we mention the word “by” we are talking about adding another dimension to the analysis. In this case, the sales manager needs only to pull the data from specific dimensions, only those which are necessary for his analysis requirements.

We may represent a set of data with predefined dimensions as a cube which closely resembles a Rubix cube. Each face of the cube represents a particular dimension; the dimensions being coexistent with each other at every point and, at the same time, are also independent from each other. Similar to twisting a Rubix cube, a user can also manipulate a multidimensional cube in several ways to view data. The key element here is that regardless of the way the data is viewed or displayed, the same intersections of data will still show the same measure, meaning the actual data is preserved.

In some multidimensional database applications such as an OLAP product, formulas are defined inside the cube, or set of dimensions. Formulas assign an equation to be performed on a specific portion of a cube, which is much faster and more effective than defining a formula in a spreadsheet application based on a specific cell. By applying the formulas directly to the data in the cube, the multidimensional database application eliminates the redundancy of redefining formulas in Excel, for example. And, heedless of the way the data is being viewed or organized, the formula will still apply to the same portion of the cube.

One technique for manipulating the data view is Rotation or more commonly known as Slicing and dicing. Slicing and dicing is a function that enables users to see the different faces of a cube in a single view by displaying them in a flat spreadsheet-like format. The reason for employing a

spreadsheet-like view of data is that a computer screen is only limited to providing a flat layout of data. This permits users to customize the output of data by moving the dimensions from rows to columns depending on a specific layout preferred by user.

Although data is presented in a two-dimensional or flat layout, it doesn't necessarily imply that only two dimensions can be displayed at once. Actually, it is very possible that there are more dimensions present in the data view. Some dimensions may simply be nested inside another dimension. How is this possible? To explain it more clearly, nesting dimensions is merely combining the intersections of dimensions to show more than two dimensions of information in a single view. This is done by nesting the dimensions either in rows, columns or both.

		2005	2006	2007
Sales	Branch 1	345,798.00	354,875.00	349,212.00
	Branch 2	597,545.00	645,231.00	591,311.00
	Branch 3	248,846.00	281,346.00	346,441.00
Cost	Branch 1	122,554.00	124,112.00	122,995.00
	Branch 2	264,465.00	284,643.00	264,111.00
	Branch 3	97,883.00	102,315.00	135,422.00
Profit	Branch 1	223,244.00	230,763.00	226,217.00
	Branch 2	333,080.00	360,588.00	327,200.00
	Branch 3	150,963.00	179,031.00	211,019.00

For example a user may want to see comparatively the Sales, Cost and Profit results of different branches for the past three years. Here we are dealing with the dimensions Year, Accounts (Aggregation of members Sales, Cost and Profit), Branch, and Measure. Rather than presenting the needed information using three different reports, the user can nest one dimension under another. For example Branch may be nested under Accounts dimension along the rows and display year along the columns.

What happens is that Accounts and Branch dimension will be displayed simultaneously on the rows. This enables the user to directly compare the information for the past three years by looking at a single data only. This is an optimized view for analysis purposes.

DATA VISUALIZATION

Visual technologies make pictures worth a thousand numbers, and decision support applications more attractive and understandable to users. Data visualization refers to technologies that support visualization and sometimes interpretation of data and information at several points along the data processing chain. It includes digital images, geographic information systems, graphical user interfaces, multi dimensions, tables and graphs, virtual reality, three-dimensional presentations, and animation. Visual tools can help identify relationships directly. The ability to identify

important trends in corporate and market data provides enormous advantages. More accurate predictive models provide significant business advantages in applications that drive content, transactions, or processes. Confident action, based on superior methods of visual data analysis, helps companies improve income and avoid costly mistakes. For example, network monitoring systems continue to become increasingly complicated and sophisticated. Visualization simplifies the reporting of test results.

For further details: refer to "***Visualisation and data issues.pdf***" shared separately

GEOGRAPHICAL INFORMATION SYSTEM

Geographic information system, (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographically referenced data. Or in simpler terms, GIS is the combination of digital data in the form of map, Statistical analysis of data a database technology.

GIS provides a framework for planners to design things more efficiently by designing capabilities. Thus, GIS integrates hardware, software, and data for capturing, managing, analyzing, and displaying all forms of geographically referenced information.GIS allows us to view, understand, question, interpret, and visualize data in many ways that reveal relationships, patterns, and trends in the form of maps, globes, reports, and charts.

GIS technology helps to answer questions and solve problems by looking at the spatial data on real time basis in a way that is quickly understood and easily shared and displays information ensuring faster decision making.

Geographic Information System (GIS) continues to accelerate, and is compelling right now, because maps communicate better than words, spatial analytics helps us understand and make decisions about our world with a broader sense and cloud computing extends further more in the line. GIS is helping to unlock in environmental conservation, ecological science, natural resources, disaster response, business analytics, crime mapping, etc. GIS is becoming societal infrastructures as a national information system. GIS no doubt massively could go beyond just simply desktop use. The issue of climate change is going to have a major impact, and GIS has a huge role to play in the desperate race to turn around the impact. GIS is evolving with many forces that are converging, including more measurement, more ubiquitous computer networks, open data policies, better geographic science, and it all feeds collective geographic understanding. GIS changes how we reason by bringing together different disciplines and creating integrated thinking.

What is the Importance of GIS Data?

Geographic Information Systems (GIS) have various industrial applications, and technological advancements have significantly enhanced GIS data, specifically how it can be used and what can be achieved as a result.

Geographic Information Systems are powerful decision-making tools for any business or industry since it allows the analysis of environmental, demographic, and topographic data. Data intelligence compiled from GIS applications help companies and various industries, and consumers, make informed decisions.

Here are 20 ways GIS Data is used in Business and Everyday Life:

1. Mapping

GIS can be used to provide a visual interpretation of data. Google Maps is an excellent example of a web-based GIS mapping solution that people use for everyday navigation purposes. However, smart mapping technology has significantly advanced and is used in products like Nobel's GeoViewer, which gives cities, municipalities and private industry an in-depth look at electric and water district assets in the field.

2. Telecom and Network Services

Organizations can incorporate geographic data into their complex network design, optimization, planning, and maintenance activities. This data enhances telecom processes through better customer-relationship management and location services.

3. Accident Analysis and Hot Spot Analysis

GIS data helps to identify accident locations, and road networks can be optimized using data intelligence. This intelligence helps to improve road safety measures and allows better traffic management.

4. Urban planning

GIS data analyzes urban growth and the direction of expansion. When appropriately applied, it can discover new sites for further development, considering various factors that are necessary for successful building.

5. Transportation Planning

GIS data is commonly used for managing transportation issues. With the addition of environmental and topical data in a GIS platform, companies can plan for a new road or rail route.

6. Environmental Impact Analysis

Data gathered via GIS applications is vital for conserving natural resources and protecting the environment. Impact statements assess the magnitude of human impact on the environment, which GIS integration helps indicate.

7. Agricultural Applications

GIS data helps create more efficient farming techniques, alongside analyzing soil data in an advanced fashion. This can increase food production in different parts of the world.

8. Disaster Management and Mitigation

Efficient GIS systems protect the environment and are developed to assist risk and disaster management.

9. Navigation

Web-based navigation maps use GIS data to provide the public with useful information. Web maps are regularly updated per GIS information and are used consistently in everyday life.

10. Flood damage estimation

Governments use GIS data to map flood risk areas and can use the information to coordinate relief efforts.

11. Natural Resources Management

With the help of GIS information, forests can be adequately maintained and managed. It is especially crucial for the allocation and geographic distribution of water, one of the more critical environmental constituents.

12. Banking

Banking has evolved to become market-driven, and a bank's success depends mainly on its ability to provide customer-driven services. GIS data plays an essential role in planning, organizing, and decision making in the banking industry.

13. Taxation

GIS data helps solve taxation problems and maximize government income. It is used for building permits and engineering and offers a system for managing property tax on a geographic basis.

14. Surveying

Surveying involves measuring the location of objects on earth, and more organizations are using Global Navigation Satellite Systems (GNSS) for this function. This data incorporated into a GIS system can estimate area and prepare digital maps.

15. Geology

Geologists use GIS data to analyze soil, assess seismic information, and create 3D displays of geographic features. It can also be used to analyze rock characteristics, and identify the best location for different functions.

16. Assets Management and Maintenance

GIS data helps organizations become more efficient with finite resources. With an understanding of the population at risk, planners can allocate resources more efficiently.

17. Planning and Community Development

GIS data helps us understand and meet global challenges. As GIS technology rapidly advances, there are various innovative applications in the planning sector. GIS tools can be used to integrate geographic intelligence into planning processes, and have the potential to change how we think and behave.

18. Dairy Industry

The dairy industry uses GIS data for distribution, production, and identifying the location of shops. It is a useful tool for planning in the field of dairy farm management and allows for better decision making.

19. Irrigation Water Management

The availability of water directly affects crop production in a given region. GIS data can identify significant crops and determine yield, involving efficient techniques for spatial and time domain.

20. Pest Control and Management

Pest control is essential to agricultural production, and GIS technology plays a vital role in mapping out infested areas. Organizations can consequently develop more effective pest management plans.

Real time Business Intelligence Automated Decision support

Businesses these days are exploring innovative ways to transform their operations effectively. They are increasingly tapping into real-time analytics to respond quickly to customers and deliver a variety of services to them. Unlike conventional data analysis, real-time analytics uses and analyzes the data as soon as it enters in the computer system. This helps organizations to gain a competitive edge and improve agility. Real-time analytics often takes place at the edge of the network ensuring that data analysis is done as close to the origin of data as possible.

Real-time analytics plays a crucial role in the acceptance rates and life-cycle development of Big Data. As it dramatically transforms the ways systems use data to envisage outcomes and suggest alternatives, companies are widely turning to this to drive innovation.

Here are few companies using Real-Time Analytics to enhance business efficiency.

- **Amazon**

E-commerce giant Amazon is one of the companies enabling data-driven culture within the organization. The company gleans over 2,000 historical and real-time data points on every order and leverages machine learning algorithms to find transactions with an elevated likelihood of being fraudulent. By doing so, the company's system stops millions of dollars worth of fraudulent transactions each year. Amazon uses Big Data to automatically customize the browsing experience for its customers based on their past purchases and optimize sales.

- **Maruti Suzuki**

Another example is India's biggest automobile company – Maruti Suzuki, which is leveraging Qlik to analyse and provide business insights and improve operational efficiency across the enterprise. With Qlik's solution, Maruti Suzuki has been successful in streamlining its sales and distribution management. It has also been able to monitor supplier and dealer performance more effectively; manage complex after-sales and service opportunities efficiently and improve the process of spare parts management.

- **Nissan Motor**

Automaker giant Nissan uses Google analytics e-commerce tracking to amass detailed information about product preferences such as car category, model, and color. By assessing this information, the company's Global Marketing Strategy division understands which vehicles are in demand, thus they can make decisions tailored for each local market. The auto company has a host of localized websites aimed at assisting consumers to determine which Nissan product is best for them. Nissan has also deployed the Hortonworks Data Platform (HDP) to power its data lake. The company developed its data lake infrastructure using Apache Hadoop powered HDP to gather all data from across the business, including driving data and quality data.

- **Shell**

Shell, a Netherlands-based oil and gas company, developed an analytics platform based on software from several vendors to run predictive models to foresee when its different oil drilling machine parts might fail. The company used Databricks that captures streaming data through Apache Spark, to better plan when to purchase machine parts, how long to keep them, and where to place inventory items. Hosted in Microsoft Azure's cloud, the tool helped Shell by reducing inventory analysis from more than 48 hours to less than 45 minutes, cutting off millions of dollars a year of moving and reallocating inventory.

COMPETITIVE INTELLIGENCE

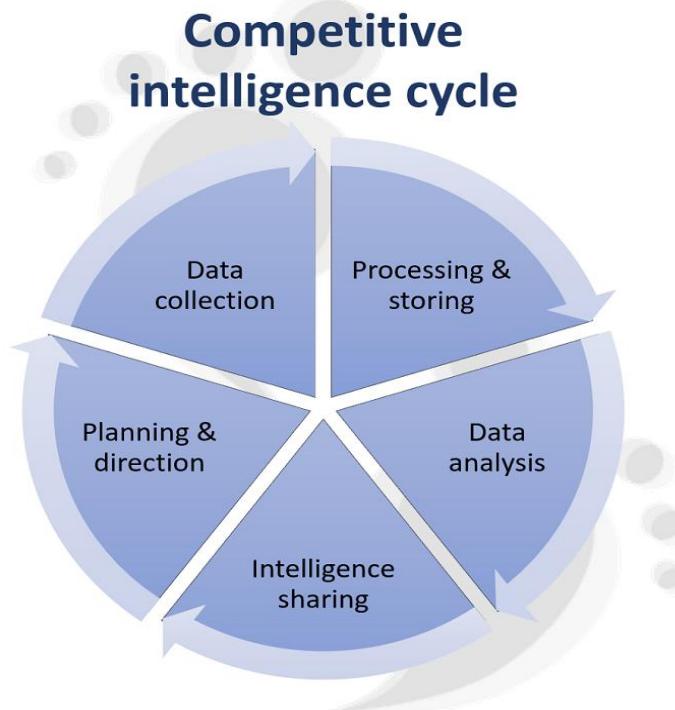
Competitive intelligence is the systematic process of observing, collecting and analyzing relevant information about the external business environment and distributing the resulting insights within an organization in order to make informed decisions. It can focus on competitors, customers and other stakeholders, but also on products and markets as well as on economic, technological or legal factors. Regardless of the scope, its main purpose is to understand the external environment so that appropriate actions can be taken by businesses to stay competitive.

It's a way to improve the competitiveness of your business, relative to its entire domain - competitors, technologies, customers, distributors, and macroeconomic data.

Information => intelligence => action

- Information is the starting point - stuff available to all of us - numbers, data about competitors, stats, company info, product intel, marketing strategies
- Intelligence is information that's been analyzed, with insights squeezed out to enable decision making, such as whether to discontinue a product or which audience segment to target

It's assumed that competitive intelligence and competitor analysis are the same thing. This would be an inaccurate assumption.



What are the benefits of competitive intelligence?

We're suffering from information overload. The amount of content being shared online alone, is overwhelming. Having a competitive intelligence program in place in every organization is the only way to tackle this heap of data. And, it should be used by all teams - R&D, HR, sales, PR, product, marketing, customer service - to help them :

- Find new markets to enter or increase presence
- Predict competitors' actions
- Improve product development
- Beat competitors to market

- Improve and personalize consumer interaction
- Give consumers what they want and increase market share
- Find new products and tech that will disrupt the competitive landscape
- Identify legislative or political issues
- Stand out from the crowd with a distinct corporate identity and tone of voice

Competitive intelligence examples

Airline tickets

The airline industry is a great example of how competitive intelligence is being used in practice. Every day, airline companies are changing their flight ticket prices based on several pieces of external information. For instance, if all competitors increase their price for a certain route, a flight provider would quickly follow suit to secure higher margins. In addition, customer information is frequently used for pricing adjustments. By identifying and tracking specific users, flight companies can spot when a potential customer is repeatedly searching for the same flight details and increase the prices over time, since they can be sure that they really want to fly on these dates.

Investment trading

Investment banking and trading is another perfect case for highlighting the importance of competitive intelligence. While all bankers have access to the same information (excluding illegal cases of insider trading) through news channels, financial statement reports, industry research papers etc., those who know how to use and analyze the available information to gain valuable insights will have the best chances to outperform the rest.

Tech startups

Startups are also fantastic competitive intelligence examples, as they use this kind of information to disrupt markets. Whereas traditional companies have higher budgets, resources superior technology and often data too, startups are often able to outperform them in specific niche segments. By focusing on a particular field and processing, reacting and adapting rapidly to competitive intelligence insights, they can understand client pain points better and deliver superior solutions. Think of Airbnb and how they were able to leverage technology, socio-demographic change and consumer insights to disrupt the hotel industry.

Web Analytics:

Web analytics deals with collection and analysis of data on the user visits on websites and cloud applications. Analysis of this data can give insights about the user engagement and tracking the performance of online advertisement campaigns. For collecting data on user visits, two approaches

are used. In the first approach, user visits are logged on the web server which collects data such as the date and time of visit, resource requested, user's IP address, HTTP status code, for instance. The second approach, called page tagging, uses a JavaScript which is embedded in the web page. Whenever a user visits a web page, the JavaScript collects user data and sends it to a third party data collection server. A cookie is assigned to the user which identifies the user during the visit and the subsequent visits. The benefit of the page tagging approach is that it facilitates real-time data collection and analysis. This approach allows third party services, which do not have access to the web server (serving the website) to collect and process the data. These specialized analytics service providers (such as Google Analytics) offer advanced analytics and summarized reports. The key reporting metrics include user sessions, page visits, top entry and exit pages, bounce rate, most visited page, time spent on each page, number of unique visitors, number of repeat visitors, for instance.

Web analytics is the process of analyzing the behavior of visitors to a website. This involves tracking, reviewing and reporting data to measure web activity, including the use of a website and its components, such as webpages, images and videos.

Data collected through web analytics may include traffic sources, referring sites, page views, paths taken and conversion rates. The compiled data often forms a part of customer relationship management analytics (CRM analytics) to facilitate and streamline better business decisions.

Web analytics enables a business to retain customers, attract more visitors and increase the dollar volume each customer spends.

Analytics can help in the following ways:

- Determine the likelihood that a given customer will repurchase a product after purchasing it in the past.
- Personalize the site to customers who visit it repeatedly.
- Monitor the amount of money individual customers or specific groups of customers spend.
- Observe the geographic regions from which the most and the least customers visit the site and purchase specific products.
- Predict which products customers are most and least likely to buy in the future.

The objective of web analytics is to serve as a business metric for promoting specific products to the customers who are most likely to buy them and to determine which products a specific customer is most likely to purchase. This can help improve the ratio of revenue to marketing costs.

In addition to these features, web analytics may track the clickthrough and drilldown behavior of customers within a website, determine the sites from which customers most often arrive, and communicate with browsers to track and analyze online behavior. The results of web analytics are provided in the form of tables, charts and graphs.



Web analytics process

The web analytics process involves the following steps:

- Setting goals.** The first step in the web analytics process is for businesses to determine goals and the end results they are trying to achieve. These goals can include increased sales, customer satisfaction and brand awareness. Business goals can be both quantitative and qualitative.
- Collecting data.** The second step in web analytics is the collection and storage of data. Businesses can collect data directly from a website or web analytics tool, such as Google Analytics. The data mainly comes from Hypertext Transfer Protocol requests -- including data at the network and application levels -- and can be combined with external data to interpret web usage. For example, a user's Internet Protocol address is typically associated with many factors, including geographic location and clickthrough rates.
- Processing data.** The next stage of the web analytics funnel involves businesses processing the collected data into actionable information.
- Identifying key performance indicators (KPIs).** In web analytics, a KPI is a quantifiable measure to monitor and analyze user behavior on a website. Examples include bounce rates, unique users, user sessions and on-site search queries.
- Developing a strategy.** This stage involves implementing insights to formulate strategies that align with an organization's goals. For example, search queries conducted on-site can help an organization develop a content strategy based on what users are searching for on its website.
- Experimenting and testing.** Businesses need to experiment with different strategies in order to find the one that yields the best results. For example, A/B testing is a simple strategy to help learn how an audience responds to different content. The process involves creating two or more versions of content and then displaying it to different audience segments to reveal which version of the content performs better.

The two main categories of web analytics are off-site web analytics and on-site web analytics.

Off-site web analytics

The term off-site web analytics refers to the practice of monitoring visitor activity outside of an organization's website to measure potential audience. Off-site web analytics provides an industrywide analysis that gives insight into how a business is performing in comparison to competitors. It refers to the type of analytics that focuses on data collected from across the web, such as social media, search engines and forums.

On-site web analytics

On-site web analytics refers to a narrower focus that uses analytics to track the activity of visitors to a specific site to see how the site is performing. The data gathered is usually more relevant to a site's owner and can include details on site engagement, such as what content is most popular. Two technological approaches to on-site web analytics include log file analysis and page tagging.

Log file analysis, also known as log management, is the process of analyzing data gathered from log files to monitor, troubleshoot and report on the performance of a website. Log files hold records of virtually every action taken on a network server, such as a web server, email server, database server or file server.

Page tagging is the process of adding snippets of code into a website's HyperText Markup Language code using a tag management system to track website visitors and their interactions across the website. These snippets of code are called tags. When businesses add these tags to a website, they can be used to track any number of metrics, such as the number of pages viewed, the number of unique visitors and the number of specific products viewed.

Web analytics tools

Web analytics tools report important statistics on a website, such as where visitors came from, how long they stayed, how they found the site and their online activity while on the site. In addition to web analytics, these tools are commonly used for product analytics, social media analytics and marketing analytics.

Business Analytics vs. Web Analytics

Web analytics and business analytics may sound similar, but they actually have very different functions. This chart breaks down the basic differences:

	Business Analytics	Web Analytics
Analyzes Business Data Collects data on KPIs from databases or other proprietary sources and searches for trends and patterns		
Analyzes Website Data Collects data from website such as traffic, clicks, country of origin, etc.		
Presents Data in Intuitive Visualizations Organizes data into dashboards, graphs and other visualizations so users can easily interpret patterns		
Optimizes Website Helps users formulate a web strategy to improve your website's effectiveness		
Optimizes Business Helps users strategize business practice improvements based on data		

source: SelectHub

Web Analytics and Business Intelligence

If we take this definition and apply it to a company department which focuses on the company's online activity, and then add to it the data generated by the company's information systems (CRM, databases, marketing items etc.), it is clear that **Web Analytics** is Business Intelligence.

The difference between the two exists not in the wording that is used but in the thoughts of the specialists in each domain. Business Intelligence is created by large technological solutions such as the cube, distributed by the leading names in software. The main difference is in the complexity of Business Intelligence type projects, which are generally led by chief information officers and in a specifically targeted market: a market of experts.

Web Analytics has led to the creation of a new type of "key in hand" solution which is much more "work" oriented than Business Intelligence. By default Business Intelligence is somewhat

generalist, lacking in consistency and does not have any standards. Only clients are able to determine their own needs and the way in which information is to be collected, as well as the different sources, the quality of data, rules to be applied etc, right through to the KPIs which are then to be shown on a dashboard, which also needs to be designed. Business Intelligence projects are often large, long and expensive projects, which rely on the use of complex solutions.

This is not the case for Web Analytics, however. A large part of the data which is collected is externalised and standardised, including variables which can be compared with one another, standards set by the WAA* (or OJD in France, the ABCe in the UK etc.), processing systems which are specific to each solution (managed by editors and easy to understand by the client), as well as an intuitive web interface made possible by this type of support and which can be used by all services within a company and not just for an elite few.

Business Intelligence is a method which can be used; Web Analytics on the other hand is a solution.

There may, however, be a significant gap in the field of Web Analytics: the quality of the database engine and data reproduction should make it possible to carry out advanced analyses and all types of cross analyses, whilst at the same time remaining accessible to the final user. This is the point where several Web Analytics solutions remain reporting solutions, and have difficulty in becoming Business Intelligence solutions. AT Internet is different from its competitors with its complete Analyzer package associated with the **Data Explorer module**, an advanced query decision engine.

This is where Online Intelligence differs as the agility from Web Analytics is combined with the processing power associated with the field of Business Intelligence.

Faced with the increasing complexity of online projects, is Web Analytics going to finally transform into Business Intelligence? It would be somewhat presumptuous to pretend to have the response, however, it would be a shame to turn a practical, quick, and simple yet powerful solution into a Business Intelligence tool, as is often the case. The Business Intelligence solutions which are currently in place are not adapted to the world of the Internet, the web or mobile. This is a world which is continually advancing, where it is necessary to constantly adapt to new features, and a world which requires reactivity.

AT Internet's promise is, more than ever, to increase its capabilities and the powerful analyses of its solutions, whilst at the same time holding onto the notion of simplicity, clarity and agility which are required from the tagging process right through to analysing data. This is an important challenge because what is at stake is the creation of a new business intelligence tool which is more accessible than ever before.

Unit II- VISUALIZATION AND DATA ISSUES

ORGANIZATION OF SOURCE OF DATA

Data is raw material for data processing. Data relates to fact, event and transactions. Data refers to unprocessed information. It is the lowest abstract or a raw input which when processed or arranged makes meaningful output.

Sources of Data

There are two sources of data in Statistics. Statistical sources refer to data that are collected for some official purposes and include censuses and officially conducted surveys. Non-statistical sources refer to the data that are collected for other administrative purposes or for the private sector.

- Statistical Survey**

A statistical Survey is normally conducted using a sample. It is also called Sample Survey. It is the method of collecting sample data and analyzing it using statistical methods. This is done to make estimations about population characteristics. The advantage is that it gives you full control over the data. One can ask questions suited to the study you are carrying out. But the disadvantage is that there is a chance of sample error creeping up. This is because a sample is chosen and the entire population is not studied. Leaving out some units of the population while choosing the sample causes this error to arise.

- Census**

Opposite to a sample survey, a census is based on all items of the population and then data are analyzed. Data collection happens for a specific reference period. For example, the Census of India is conducted every 10 years. Other censuses are conducted roughly every 5-10 years. Data is collected using questionnaires that may be mailed to the respondents.

Responses can also be collected over other modes of communication like the telephone. An advantage is that even the most remote of the units of the population get included in the census method. The major disadvantage lies in the high cost of data collection and that it is a timeconsuming process.

- Register**

Registers are basically storehouses of statistical information from which data can be collected and analysis can be made. Registers tend to be detailed and extensive. It is beneficial to use data from here as it is reliable. Two or more registers can be linked together based on common information for even more relevant data collection.

From agriculture to business, all industries maintain registers for record-keeping. Some administrative registers also serve the purpose of acting as a repository of data for other statistical

bodies in a country.

Information is data that has been processed in such a way as to be meaningful to the person who receives it. It is anything that is communicated. Information can be explained as any kind of understanding or knowledge that can be exchanged with people. It can be about facts, things, concepts, or anything relevant to the topic concerned. **for example:** Researchers who conduct market research survey might ask member of the public to complete questionnaires about a product or a service. These completed questionnaires are data; they are processed and analyze in order to prepare a report on the survey. This resulting report is information.

Types of Information

Depending on the various sources the information is collected, we can divide information into three types:

Primary

These are original materials which have not been filtered through interpretation, condensation, or, often, even evaluation by a second party; for example journal articles, monographs, reports, patents, theses, diaries, letters, photographs, poems.

Secondary Sources

A secondary source is information about primary, or original, information, which usually has been modified, selected, or rearranged for a specific purpose or audience. It is not always easy to discern the difference between primary and secondary sources. Examples include biographies, histories, monographs, review articles, textbooks, and any index or bibliography used to locate primary sources.

Tertiary Sources

These consist of information, which is a distillation and collection of primary and secondary sources. Twice removed from the original, they include encyclopedias, fact books and almanacs, guides and handbooks. Some secondary sources such as indexing and abstracting tools can also be considered tertiary sources.

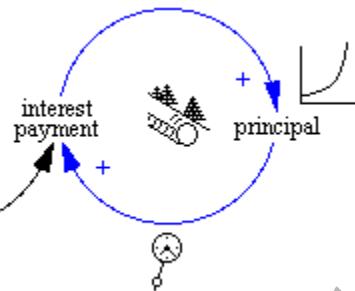
Data: The numbers 100 or 5%, completely out of context, are just pieces of data. Interest, principal, and interest rate, out of context, are not much more than data as each has multiple meanings which are context dependent.

Information: If I establish a bank savings account as the basis for context, then interest, principal, and interest rate become meaningful in that context with specific interpretations.

Principal is the amount of money, Rs.100, in the savings account.

Interest rate, 5%, is the factor used by the bank to compute interest on the principal.

Knowledge: If I put Rs.100 in my savings account, and the bank pays 5% interest yearly, then at the end of one year the bank will compute the interest of Rs.5 and add it to principal and I will have Rs.105 in the bank. This pattern represents knowledge, which, when I understand it, allows me to understand how the pattern will evolve over time and the results it will produce. In understanding the pattern, I know, and what I know is knowledge. If I deposit more money in my account, I will earn more interest, while if I withdraw money from my account, I will earn less interest.

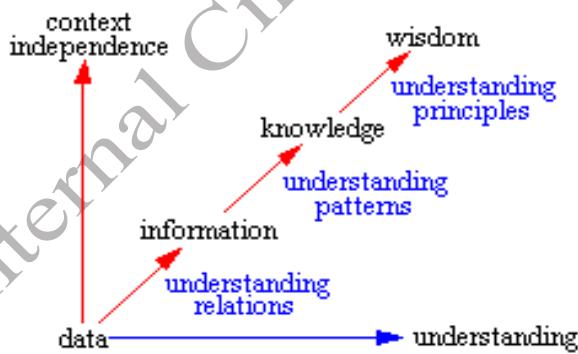


Wisdom: Getting wisdom out of this is a bit tricky, and is, in fact, founded in systems principles. The principle is that any action which produces a result which encourages more of the same action produces an emergent characteristic called growth. And, nothing grows forever for sooner or later growth runs into limits.

If one studied all the individual components of this pattern, which represents knowledge, they would never discover the emergent characteristic of growth. Only when the pattern connects, interacts, and evolves over time, does the principle exhibit the characteristic of growth.

An Example

This example uses a bank savings account to show how data, information, knowledge, and wisdom relate to principal, interest rate, and interest.



In this age of information, knowledge is the most important factor in the long-term success of both an individual and an organization. In fact, some people believe that the only source of competitive advantage in the future will be the knowledge that an organization contains and an organization's ability to learn faster than the competition

When all the people in an organization understand what they already know, what they need to know and what they don't yet know about something but need to find out, then people can work together better to reach a goal. Individuals will get more accomplished because of the knowledge shared within a group, and groups can collaborate better because they have a common goal and a clear starting point.

The benefits of knowledge management can take a company that's struggling with the information processing cycle and productivity and make it a much more efficient and success-oriented workplace.

Quality and Value of Information

Today, information quality is understood to be a multidimensional concept that encompasses critical relationships among multiple attributes, such as timeliness, accuracy, relevancy, and others. Together, these attributes contribute to the validity of the information. Quality information is the cornerstone of sound agency decision making.

Three dimensions of that help in assessing quality of Information are:

TIME

- Timeliness**

It means that every information is helpful and is considered to be of good quality only if it is provided timely.

Eg: If newspapers provide news about events after two days of its occurrence, it can easily loose its importance.

- Currency**

A good quality of information is one which gives fact about the current status and provides the position it is in that very moment.

Eg: in hospital, patient is treated on the basis of his/her current condition and not on the basis of condition and health he had some time back.

- Frequency**

Some of the information requires that it should be provided on a specific frequency else it loses its relevance. This frequency can be every minute, hour, daily, monthly, etc depending upon the need.

Eg: In case of a cricket match, we need news about what happened at every ball and if its provided only after an hour, such information loses its importance

- Time Period**

Whenever any information is provided, it always depicts a particular time period which increases or decreases its relevance.

Eg: During excavation, any hidden or buried article when extracted tells about the time period to which it belongs. If the article obtained is not in a position to give such information, it becomes useless.

CONTENT

- **Accuracy**

The information provided should always be accurate in nature. Inaccurate information might make any useful information, useless and also prove to be fatal or dangerous.

Eg: While going for eye check-up, the *ophthalmologist* is expected to provide accurate results else wrong numbers can be fatal for the patient.

- **Relevance**

If the provided information is of no relevance at the given time, it loses its importance.

Eg: News about a possible terror attack provided after it has already occurred is of importance.

- **Completeness**

Any information if is provided incomplete can lose its importance.

Eg: Information related to possible traffic jams in a city is incomplete till the time it is not provided along with the exact location and if possible time period.

- **Conciseness**

Any information expressed using few words which are brief and to the point without providing unnecessary and irrelevant information.

Eg: Any dictionary which only provides meaning and not the history and how to use it is often useful for providing just the right information.

- **Scope**

Scope of coverage refers to the extent to which a source explores a topic. Consider time periods, geography or jurisdiction and coverage of related or narrower topics.

Eg: People acquire education as it increases the ability and provides scope to get a good job.

- **Performance**

A good quality information is capable of enhancing the performance of the person who obtains it.

Eg: Subjects taught in classrooms provide information keeping in mind it increases the mental ability of the students. If it fails to do so, such information is useless.

FORM

- **Clarity**

A good information has clarity of thoughts and also has to be clear in the way it is expressed.

Eg: The announcement about airlines , if unclear to the passengers, proves to be useless as it is not capable of providing the passengers with any assistance.

- **Detail**

Some of the information needs to be provided in detail else it loses its importance.

Eg: If subjects taught in schools do not provide complete information in detail, students are unable to understand the basic concept.

- **Order**

The information needs to be in proper order. If they are arranged in haphazard manner, it loses its relevance and demands time to understand it.

Eg: Information provided in a resume of a candidate is supposed to be in order and is expected not to be provided one after the other.

- **Presentation**

The way information is presented can increase or decrease its importance.

Eg: If a project made with lots of hard work is finally presented in loose papers, in no order, it can lose its importance.

- **Media**

The way information is exchanged can also determine or undermine its importance.

Eg: Newspapers have important news and lines presented as headlines, in different font sizes in order to attract attention to the important information.

Mastering Data Quality

The data quality issue is an orphan in many Business Intelligence (BI) projects. It is often addressed when the embarrassing results of the first reports have come to the attention of the business sponsor. One of the reasons is the lack of understanding what data quality for Business Intelligence implies. From a transaction point of view, 99% of the data may be correct but from an analyst's point of view these same data may be useless. So how do we address the issue before the data warehouse is built and before the business sponsor becomes annoyed?

When we are talking data quality, what are we really talking about? Is it product quality? Is it user-defined quality? Production-oriented data quality? Value-based quality or even transcendental quality? The quality perspective for data quality is Joseph Juran's (2010) "fitness-for-use" approach. Data quality for BI purposes is defined and gauged with reference to fitness for

purpose as defined by the analytical use of the data and complying with three levels of data quality, as defined by:

- [Level 1] database administrators
- [Level 2] data warehouse architects
- [Level 3] business intelligence analysts

On level 1, data quality is narrowed down to data integrity or the degree to which the attributes of an instance describe the instance accurately and whether the attributes are valid, that is, comply with defined ranges or definitions managed by the business users. This definition remains very close to the transaction view.

On level 2, data quality is expressed as the percentage completeness and correctness of the analytical perspectives. In other words, to what degree is each dimension, each fact table, complete enough to produce significant information for analytical purposes? Issues such as sparsity and spreads in the data values are harder to tackle. Timeliness and consistency need to be controlled and managed on the data warehouse level.

On level 3, data quality is the measure in which the available data are capable of adequately answering the business questions. Some use the criterion of accessibility with regard to the usability and clarity of the data. Although this seems a somewhat vague definition, it is most relevant to anyone with some analytical mileage on his odometer. I remember a vast data-mining project in a mail-order company producing the following astonishing result: 99.9% of all dresses sold were bought by women!

Although there is no 100% data quality possible on this planet and although we defined the fit-for-purpose quality approach as the leading criterion, this does not dismiss us from striving toward the optimum solution, namely the breakeven point between data quality prevention costs and the cost of poor data quality.

Difference between Information and Intelligence

Difference between information and intelligence is distinct in their definitions and meanings, but it is an interesting topic to discuss as they are two interrelated subjects. Both terms, information and intelligence, function as nouns in English language. Information is data or knowledge of something learned or gained from somewhere. Intelligence, on the other hand, can be defined as the ability to comprehend, understand, logic, memory planning, etc. Information is available everywhere to anybody, but intelligence could be different from one person to another. However, both terms have an interrelationship with each other. Information may help to increase the level of intelligence in a person.

What does Information mean?

Information is data that contain a message or knowledge of something and also it can be shortened as “info” as well. Everything a person knows can be regarded as knowledge and this knowledge is usually in the form of information. Information can give answers to problems that arise in humans, since they carry knowledge. It can be seen that somebody requires his/her

intelligence to acquire information. Information may not come to a person, but the person may have to search for it. Thus, it needs to have intelligence. In one's process of education, he/she collects information and expands their existing knowledge.

There are various methods through which a person can get information on different things. By reading, observing, talking with each other, researching one can have access to information. Also, information can be encoded to different media and they can be transmitted through speech, signals or signs. Information interpretation, however, requires knowledge and abilities of a particular person who receives information.

What does Intelligence mean?

Intelligence can be defined as the intellectual capacity of a human being or any other species. As mentioned above, intelligence is one of the main requirements to have access to different information. Intelligence is usually characterized by the ability to perceive something, understanding, logical thinking and self-awareness, etc. Due to intelligence, human beings gain the cognitive ability to learn and to analyze various things. Moreover, intelligence is the driving force of humans to solve their problems with reasoning, to plan certain things and most importantly to use a language to communicate and share ideas. It is because of the intelligence that any person gets to experience the material and conceptual world around them. It enables individuals' thinking and experiencing power of their surroundings. However, intelligence is not equal in each and every human being. Due to many factors, the level of intelligence may be varied from one person to another. It is important to note that, it is not only humans who possess intelligence. Even animal beings have their own intellectual capacities. Further, now we have robots or artificial intelligence, created by humans, who also have a considerable level of intelligence.

What is the difference between Information and Intelligence?

When we consider about both terms, we see an interrelationship between them.

- Information is available to any person, anywhere in the world equally.
- In contrast, intelligence is something innate to humans and the level of intelligence varies from one person to another.
- Information collection of a person depends on the level of that person's intelligence. In that sense, there is an interrelationship between the two terms.
- However, both information and intelligence are important to humans because they survive their lives based on these two factors. When a problem arises, both information and intelligence are needed to find a solution.

Information Asymmetry

Information asymmetry deals with the study of decisions in transactions where one party has more or better information than the other. Information asymmetry is an imbalance between two negotiating parties in their knowledge of relevant factors and details. Typically, that imbalance means that the side with more information enjoys a competitive advantage over the other party.

Asymmetric information, sometimes referred to as information failure, is the specialization and division of knowledge in society. For example, medical doctors tend to know more about medical treatment than their patients; after all, those doctors specialize in medicine, while their patients do not. The same principle applies to manufacturers, teachers, police officers, attorneys, restaurant operators and yoga instructors, or any other specialized profession. A common assumption in researching such issues is that managers will make decisions that perfectly reveal the firm's prospects to the less-informed party, even if it is costly to do so. For example, a firm facing a big market opportunity will open more stores than is optimal in order to signal its favorable prospects. The number of stores the firm opens must be so high that a firm facing a small opportunity will find it too expensive to mimic the number of store openings. While such predicted outcomes underpin much of the operations theory developed in these settings, they have not been reconciled against the decisions made by actual decision makers.

Dealing with Missing or Incomplete Data

Missing values are the Achilles's heel for a data scientist. If not handled properly, the entire analysis will be futile and provide misleading results which could potentially harm the business stakeholders.

Missing data, where either entire observations or individual variable values are for some reason not available for analysis, is a common challenge to research using complex data bases.

Although missing data clearly lead to a loss of information and hence reduced statistical power, a more insidious consequence is that this lack of data may introduce selection bias, which could potentially invalidate the entire study. Fortunately, many solutions to this problem have been put forth, based on different ways of filling out (imputing) values where there are none.

Unfortunately, several of these solutions are inherently flawed and may introduce more bias than they remove.

When a variable is missing due to incomplete data, ideally the variation it accounts for should sit in the residual as a whole. Post-modeling checks would then indicate violation of the random residual assumption which would imply a missing variable(s) in the regression model. However in practice, this is almost always not the case and the diagnosis and fixing of the problem of incomplete data can almost be considered an art. If the variable missing is correlated with others included in the model, then it is quite possible to encounter omitted variable biases. This could result in either overestimation or underestimation of the relationship between variables in the

model and the dependent. It is not possible to determine the exact level of bias in the models which adds an additional layer to complexity to the problem.

Types of Missing Data:

D.B Rubin (1976) classified missing data problems into three categories. In his theory every data point has some likelihood of being missing. The process that governs these probabilities is called the ‘missing data mechanism’ or ‘response mechanism’. The model for the process is called the ‘missing data model’ or ‘response model’.

Rubin’s distinction sets the conditions under which a missing data handling method can provide valid statistical inferences.

Missing Completely at Random (MCAR)

If the probability of being missing is the same for all cases, then the data are said to be missing completely at random (MCAR). This effectively implies that causes of the missing data are unrelated to the data. It is safe to ignore many of the complexities that arise because of the missing data, apart from the obvious loss of information. Most simple fixes only work under the restrictive and often unrealistic MCAR assumption.

Example: Estimate the gross annual income of a household within a certain population, which you obtain via questionnaires. In the case of MCAR, the missingness is completely random, as if some questionnaires were lost by mistake.

Missing at Random (MAR)

If the probability of being missing is the same only within groups defined by the observed data, then the data are missing at random (MAR). It is more general and more realistic than MCAR. Modern missing data methods generally start from the MAR assumption.

Example: Suppose some household income information is missing. In the case of MAR, the missingness is random within subgroups of other observed variables. For instance, suppose you also collected data on the profession of each subject in the questionnaire and deduce that managers, VIPs etc are more likely not to share their income, then, within subgroups of the profession, missingness is random.

Not Missing at Random (NMAR)

If neither MCAR nor MAR holds, then we speak of missing not at random (MNAR). In the literature one can also find the term NMAR (not missing at random) for the same concept. MNAR means that the probability of being missing varies for reasons that are unknown to us. MNAR includes the possibility that the scale produces more missing values for the heavier objects (as above), a situation that might be difficult to recognize and handle. An example of MNAR in public opinion research occurs if those with weaker opinions respond less often. MNAR is the most complex case. Strategies to handle MNAR are to find more data about the

causes for the missingness, or to perform what-if analyses to see how sensitive the results are under various scenarios.

Example: In the case of MNAR when the reason for missingness depends on the missing values itself. For instance, suppose people don't want to share their income as it is less and they are ashamed of it.

Ways to Handle Missing Values

When it comes to handling missing values, you can take the easy way or you can take the professional way.

The Easy Way:

- Ignore tuples with missing values: This approach is suitable only when the dataset is quite large and multiple values are missing within a tuple.

Is an option only if the tuples containing missing values are about 2% or less. Works with MCAR.

- Drop missing values: Only ideal if you can afford to lose a bit of data.

Is an option only if the number of missing values is 2% of the whole dataset or less. Do not use this as your first approach.

- Leave it to the algorithm: Some algorithms can factor in the missing values and learn the best imputation values for the missing data based on the training loss reduction (ie. XGBoost). Some others have the option to just ignore them (ie. LightGBM — `use_missing=false`). However, other algorithms throw an error about the missing values (ie. Scikit learn — `LinearRegression`).

Is an option only if the missing values are about 5% or less. Works with MCAR.

The Professional Way:

The drawback of dropping missing values is that you lose the entire row just for a few missing values. That is a lot of valuable data. So instead of dropping the missing values, or even ignoring them in the case of tuples, try filling in the missing values with a well calculated estimate. Professionals use two main methods of calculating missing values. They are imputation and interpolation.

Imputation

The mean or median of the other variables within a dataset. The relationship of the data need not be linear.

Types of Imputation

Note: The built-in dataset MTcars is used to demonstrate each method.

Easy Imputations

Mean/Median Imputation a.k.a Constant Values Imputation

Calculate the mean of the observed values for the variable for all individuals which are non-missing. It has the advantage of keeping the same mean and the same sample sizes.

BI Tools for PREDICTIVE ANALYTICS

It analyzes current and historical data to make predictions about future risks and opportunities. An example of this is credit scoring, which relies on an individual's current financial standing to make predictions about their future credit behavior.

What is a Data Mining Tool?

A Data mining tool is a software application that is used to discover patterns and trends from large sets of data and transform those data into more refined information. It helps you to identify unsuspected relationships amongst the data for business growth. It also allows you to analyze, simulate, plan and predict data using a single platform.

Data Mining Tools

All that AI and Machine learning inference must have got you into wondering that for data mining implementation, you'd require nothing less. That might not entirely be true, as, with the help of most straightforward databases, you can get the job done with equal accuracy.

Why Data Mining is Important?

Data mining allows you to transform raw data into useful information in order to increase business growth. It helps businesses to identify patterns and trends among their customers to drive more sales by developing a better marketing strategy and decreasing costs.

How Data Mining Works?

The Data Mining Process works in the following stages:

- Business Understanding
- Data Understanding
- Data Preparation
- Data Transformation
- Modelling
- Evaluation
- Deployment

XLMiner : <https://www.xlminer.com/>

Frontline Systems, developer of the Solver in desktop Microsoft Excel, has released XLMiner.com, a SaaS (software as a service) platform for data mining, text mining, forecasting,

and predictive analytics using only a browser. XLMiner for the Web offers business analysts point-and-click tools to create predictive analytics models themselves, without being expert data scientists or programmers.



“Rapidly growing use of advanced analytics by spreadsheet-savvy business analysts, both desktop and cloud, is a ‘stealth trend’ that many industry observers have missed,” said Daniel Fylstra, Frontline’s President and CEO. “Other vendors are talking about ‘democratizing analytics’ through easier to use tools, but Frontline has been delivering on this vision for years.”

While XLMiner emphasizes ease of use, especially for first-time users of predictive analytics, on its higher-level subscription plans it can handle large datasets and very challenging problems – for example finding “best subsets” of many variables via exhaustive search in multiple linear regression. It performs operations such as clustering, principal components analysis, text analysis and latent semantic indexing, and training of ensembles of neural networks or classification and regression trees on Frontline’s RASON Server – a cloud analytics platform on Microsoft Azure that also handles challenging optimization and simulation models for Frontline’s customers.

Unit III - Data, Text & Web Mining

DATA MINING

Databases are growing in size to a stage where traditional techniques for analysis and visualization of the data are breaking down. Data mining and KDD are concerned with extracting models and patterns of interest from large databases. Data mining can be regarded as a collection of methods for drawing inferences from data. The aims of data mining and some of its methods overlap with those of classical statistics. It should be kept in mind that both data mining and statistics are not business solutions; they are just technologies. Additionally, there are still some philosophical and methodological differences between them.

DATA MINING

Data mining can be defined as the process of extracting data, analyzing it from many dimensions or perspectives, then producing a summary of the information in a useful form that identifies relationships within the data. Data mining tools, which use a variety of techniques, including neural networks, and advanced statistics to locate patterns within the data and develop hypotheses. There are two types of data mining: descriptive, which gives information about existing data; and predictive, which makes forecasts based on the data.

This field is growing rapidly, due in large part to the increasing awareness of the potential competitive business advantage of using such information. Important knowledge has been extracted from massive scientific data, as well. What is useful information depends on the application. Each record in a data warehouse full of data is useful for daily operations, as in online transaction business and traditional database queries. Data mining is concerned with extracting more global information that is generally the property of the data as a whole. Thus, the diverse goals of data mining algorithms include: clustering the data items into groups of similar items, finding an explanatory or predictive model for a target attribute in terms of other attributes, and finding frequent patterns and sub-patterns, as well as finding trends, deviations, and interesting correlations between the attributes.

A problem is first defined, then data source and analytic tool selection are undertaken to decide the best way to approach the data. This involves a wide variety of choices.

Decision trees and decision rules are frequently the basis for data mining. They utilize symbolic and interpretable representations when developing methods for classification and regression. These methods have been developed in the fields of pattern recognition, statistics, and machine learning. Symbolic solutions can provide a high degree of insight into the decision boundaries that exist in the data and the logic underlying them. This aspect makes these predictive mining techniques particularly attractive in commercial and industrial data mining applications.

Applying machine-learning methods to inductively construct models of the data at hand has also proven successful. Neural networks have been successfully applied in a wide range of supervised

and unsupervised learning applications. Neural-network methods are not commonly used for data mining tasks because they are the most likely to produce incomprehensible results and to require long training times. Some neural-network learning algorithms exist, however, that are able to produce good models without excessive training times.

In recent years, significant interest has developed in adapting numerical and analytic techniques from statistical physics to provide algorithms and estimates for good approximate solutions to hard optimization problems. Cluster analysis is an important technique in exploratory data analysis, because there is no prior knowledge of the distribution of the observed data. Partitional clustering methods, which divide the data according to natural classes present in it, have been used in a large variety of scientific disciplines and engineering applications. The goal is to find a partition of a given data set into several compact groups. Each group indicates the presence of a distinct category in the measurements.

In all data mining applications, results are considerably subject to interpretation, since it is a search for trends and correlation rather than an examination of hypotheses based on known real-world information. The possibility for spurious results is large, and there are many cases where the information developed will be of little real value for business purposes. Nonetheless, when pay dirt is struck, the results can be extremely useful.

Interest in data mining is growing, and it has recently been spotlighted by attempts to root out terrorist profiles from data stored in government computers. In a more mundane, but lucrative application, SAS uses data mining and analytics to glean insight about influencers on various topics from postings on social networks such as Twitter, Facebook, and user forums.

Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

Data mining is the analysis step of the “knowledge discovery in databases” process, or KDD.

The term is a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence, machine learning, and business intelligence.

Often the more general terms (large scale) data analysis and analytics – or, when referring to actual methods, artificial intelligence and machine learning – are more appropriate. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis),

unusual records (anomaly detection), and dependencies (association rule mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps.

BASIC REQUIREMENTS

A corporate data warehouse or departmental data mart is useless if that data cannot be put to work. One of the primary goals of all analytic tools is to develop processes that can be used by ordinary individuals in their jobs, rather than requiring advanced statistical knowledge. At the same time, the data warehouse and information gained from data mining and data analysis needs to be compatible across a wide variety of systems. For this reason, products within this arena are evolving toward ease of use and interoperability, though these have become major challenges.

For all analytic tools, it is important to keep business goals in mind, both in selecting and deploying tools and in using them. In putting these tools to use, it is helpful to look at where they fit into the decision-making processes. The five steps in decision-making can be identified as follows:

- Develop standard reports.
- Identify exceptions; unusual situations and outcomes that indicate potential problems or advantages.
- Identify causes of the exceptions.
- Develop models for possible alternatives.
- Track effectiveness.

Standard reports are the results of normal database queries that tell how the business is performing and provide details of key business factors. When exceptions occur, the details of the situation must be easily obtainable. This can be done by data mining, or by developing hypotheses and testing them using analytic tools such as OLAP. The conclusions can then be tested using "what-if" scenarios with simple tools such as spreadsheet applications. When a decision is made, and action is taken, the results must then be traced so that the decision-making process can be improved.

Although sophisticated data analysis may require the help of specialized data analysts and IT staff, the true value of these tools lies in the fact that they are coming closer to the user. The "dashboard" is becoming the leading user interface, with products such as Informatica's PowerCenter, Oracle's Hyperion Essbase, SAS Enterprise Miner and Arcplan Enterprise server tools designed to provide easily customizable personal dashboards.

One of the recurring challenges for data analysis managers is to disabuse executives and senior managers of the notion that data analysis and data mining are business panaceas. Even when the technology might promise valuable information, the cost and the time required to implement it might be prohibitive.

Data Mining Process

Before the actual data mining could occur, there are several processes involved in data mining implementation. Here's how:

Step 1: Business Research – Before you begin, you need to have a complete understanding of your enterprise's objectives, available resources, and current scenarios in alignment with its requirements. This would help create a detailed data mining plan that effectively reaches organizations' goals.

Step 2: Data Quality Checks – As the data gets collected from various sources, it needs to be checked and matched to ensure no bottlenecks in the data integration process. The quality assurance helps spot any underlying anomalies in the data, such as missing data interpolation, keeping the data in top-shape before it undergoes mining.

Step 3: Data Cleaning – It is believed that 90% of the time gets taken in the selecting, cleaning, formatting, and anonymizing data before mining.

Step 4: Data Transformation – Comprising five sub-stages, here, the processes involved make data ready into final data sets. It involves:

Data Smoothing: Here, noise is removed from the data.

Data Summary: The aggregation of data sets is applied in this process.

Data Generalization: Here, the data gets generalized by replacing any low-level data with higher-level conceptualizations.

Data Normalization: Here, data is defined in set ranges.

Data Attribute Construction: The data sets are required to be in the set of attributes before data mining.

Step 5: Data Modelling: For better identification of data patterns, several mathematical models are implemented in the dataset, based on several conditions.

Types of data that can be mined

1. Data stored in the database

A database is also called a database management system or DBMS. Every DBMS stores data that are related to each other in a way or the other. It also has a set of software programs that are used to manage data and provide easy access to it. These software programs serve a lot of purposes, including defining structure for database, making sure that the stored information remains secured and consistent, and managing different types of data access, such as shared, distributed, and concurrent.

A relational database has tables that have different names, attributes, and can store rows or records of large data sets. Every record stored in a table has a unique key. Entity-relationship model is created to provide a representation of a relational database that features entities and the relationships that exist between them.

2. Data warehouse

A data warehouse is a single data storage location that collects data from multiple sources and then stores it in the form of a unified plan. When data is stored in a data warehouse, it undergoes cleaning, integration, loading, and refreshing. Data stored in a data warehouse is organized in several parts. If you want information on data that was stored 6 or 12 months back, you will get it in the form of a summary.

3. Transactional data

Transactional database stores record that are captured as transactions. These transactions include flight booking, customer purchase, click on a website, and others. Every transaction record has a unique ID. It also lists all those items that made it a transaction.

4. Other types of data

We have a lot of other types of data as well that are known for their structure, semantic meanings, and versatility. They are used in a lot of applications. Here are a few of those data types: data streams, engineering design data, sequence data, graph data, spatial data, multimedia data, and more.

Data mining techniques



DATA MINING TECHNIQUES

1. Association:

It is one of the most used data mining techniques out of all the others. In this technique, a transaction and the relationship between its items are used to identify a pattern. This is the reason this technique is also referred to as a relation technique. It is used to conduct market basket analysis, which is done to find out all those products that customers buy together on a regular basis.

This technique is very helpful for retailers who can use it to study the buying habits of different customers. Retailers can study sales data of the past and then lookout for products that customers buy together. Then they can put those products in close proximity of each other in their retail stores to help customers save their time and to increase their sales.

2. Clustering

This technique creates meaningful object clusters that share the same characteristics. People often confuse it with classification, but if they properly understand how both these techniques work, they won't have any issue. Unlike classification that puts objects into predefined classes, clustering puts objects in classes that are defined by it.

Let us take an example. A library is full of books on different topics. Now the challenge is to organize those books in a way that readers don't have any problem in finding out books on a particular topic. We can use clustering to keep books with similarities in one shelf and then give those shelves a meaningful name. Readers looking for books on a particular topic can go straight to that shelf. They won't be required to roam the entire library to find their book.

3. Classification

This technique finds its origins in machine learning. It classifies items or variables in a data set into predefined groups or classes. It uses linear programming, statistics, decision trees, and artificial neural network in data mining, amongst other techniques. Classification is used to develop software that can be modelled in a way that it becomes capable of classifying items in a data set into different classes.

For instance, we can use it to classify all the candidates who attended an interview into two groups – the first group is the list of those candidates who were selected and the second is the list that features candidates that were rejected. Data mining software can be used to perform this classification job.

4. Prediction

This technique predicts the relationship that exists between independent and dependent variables as well as independent variables alone. It can be used to predict future profit depending on the sale. Let us assume that profit and sale are dependent and independent variables, respectively. Now, based on what the past sales data says, we can make a profit prediction of the future using a regression curve.

5. Sequential patterns

This technique aims to use transaction data, and then identify similar trends, patterns, and events in it over a period of time. The historical sales data can be used to discover items that buyers bought together at different times of the year. Business can make sense of this information by recommending customers to buy those products at times when the historical data doesn't suggest they would. Businesses can use lucrative deals and discounts to push through this recommendation.

DATA MINING APPLICATIONS

Below are some most useful data mining applications lets know more about them.

1. Healthcare- Data mining has the potential to transform the healthcare system completely. It can be used to identify best practices based on data and analytics, which can help healthcare facilities to reduce costs and improve patient outcomes. Data mining, along with machine learning, statistics, data visualization, and other techniques can be used to make a difference. It can come in handy when forecasting patients of different categories. This will help patients to receive intensive care when and where they want it. Data mining can also help healthcare insurers to identify fraudulent activities.

2. Education- Use of data mining in education is still in its nascent phase. It aims to develop techniques that can use data coming out of education environments for knowledge exploration. The purposes that these techniques are expected to serve include studying how educational support impacts students, supporting the future-leaning needs of students, and promoting the science of learning amongst others. Educational institutions can use these techniques to not only predict how

students are going to do in examinations but also make accurate decisions. With this knowledge, these institutions can focus more on their teaching pedagogy.

3. Market basket analysis - This is a modelling technique that uses hypothesis as a basis. The hypothesis says that if you purchase certain products, then it is highly likely that you will also purchase products that don't belong to that group that you usually purchase from. Retailers can use this technique to understand the buying habits of their customers. Retailers can use this information to make changes in the layout of their store and to make shopping a lot easier and less time consuming for customers.

4. Customer relationship management (CRM) - CRM involves acquiring and keeping customers, improving loyalty, and employing customer-centric strategies. Every business needs customer data to analyze it and use the findings in a way that they can build a long-lasting relationship with their customers. Data mining can help them do that.

5. Manufacturing engineering- A manufacturing company relies a lot on the data or information available to it. Data mining can help these companies in identifying patterns in processes that are too complex for a human mind to understand. They can identify the relationships that exist between different system-level designing elements, including customer data needs, architecture, and portfolio of products.

Data mining can also prove useful in forecasting the overall time required for product development, the cost involved in the process, and the expectations companies can have from the final product.

6. Finance and banking- The banking system has been witnessing the generation of massive amounts of data from the time it underwent digitalization. Bankers can use data mining techniques to solve the banking and financial problems that businesses face by finding out correlations and trends in market costs and business information. This job is too difficult without data mining as the volume of data that they are dealing with is too large. Managers in the banking and financial sectors can use this information to acquire, retain, and maintain a customer.

7. Fraud detection -Fraudulent activities cost businesses billions of dollars every year. Methods that are usually used for detecting frauds are too complex and time-consuming. Data mining provides a simple alternative. Every ideal fraud detection system needs to protect user data in all circumstances. A method is supervised to collect data, and then this data is categorized into fraudulent or non-fraudulent data. This data is used in training a model that identifies every document as fraudulent or non-fraudulent.

8. Monitoring Patterns -Known as one of the fundamental data mining techniques, it generally comprises tracking data patterns to derive business conclusions. For an organization, it could mean anything from identifying sales upsurge or tapping newer demographics.

9. Classification - To derive relevant metadata, the classification technique in data mining helps in differentiating data into separate classes:

- **Based on the type of data sources, mined**

Depending on the type of data handled like text-based data, multimedia data, spatial data, time-series data, etc.

- **Based on the data framework involved**

Any data set that is based on the object-oriented database, relational database, etc.

- **Based on data mining functionalities**

Here the data sets are differentiated based on the approach taken like Machine Learning, Algorithms, Statistics, Database or data warehouse, etc.

- **Based on user interaction in data mining**

The datasets are used to differentiate based on query-driven systems, autonomous systems.

10. Association

Otherwise known as relation technique, the data is identified based on the relationship between the values in the same transaction. It is especially handy for organizations trying to spot trends into purchases or product preferences. Since it is related to customers' shopping behavior, an organization can break down data patterns based on the buyers' purchase histories.

11. Anomaly Detection

If a data item is identified that does not match up to a precedent behavior, it is an outlier or an exception. This method digs deep into the process of the creation of such exceptions and backs it with critical information.

Generally, anomalies can be aloof in its origin, but it also comes with the possibility of finding out a focus area. Therefore, businesses often use this method to trace system intrusion, error detection, and keeping a check on the system's overall health. Experts prefer the emission of anomalies from the data sets to increase the chances of correctness.

12. Clustering

Just as it sounds, this technique involves collating identical data objects into the same clusters. Based on the dissimilarities, the groups often consist of using metrics to facilitate maximum data association. Such processes can be helpful to profile customers based on their income, shopping frequency, etc.

13. Regression

A data mining process that helps in predicting customer behavior and yield, it is used by enterprises to understand the correlation and independence of variables in an environment. For product development, such analysis can help understand the influence of factors like market demands, competition, etc.

14. Prediction

As implied in its name, this compelling data mining technique helps enterprises to match patterns based on current and historical data records for predictive analysis of the future. While some of the approaches involve Artificial Intelligence and Machine Learning aspects, some can be conducted via simple algorithms.

Organizations can often predict profits, derive regression values, and more with such data mining techniques.

15. Sequential Patterns

It is used to identify striking patterns, trends in the transaction data available in the given time. For discovering items that customers prefer to buy at different times of the year, businesses offer deals on such products.

16. Decision Trees

One of the most commonly used data mining techniques; here, a simple condition is the crux of the method. Since such terms have multiple answers, each of the solutions further branches out into more states until the conclusion is reached. Learn more about decision trees.

17. Visualization

No data is useful without visualizing the right way since it's always changing. The different colors and objects can reveal valuable trends, patterns, and insights into the vast datasets. Therefore, businesses often turn to data visualization dashboards that automate the process of generating numerical models.

18. Neural Networks

It represents the connection of a particular machine learning model to an AI-based learning technique. Since it is inspired by the neural multi-layer system found in human anatomy, it represents the working of machine learning models in precision. It can be increasingly complex and therefore needs to be dealt with extreme care.

19. Data Warehousing

While it means data storage, it symbolizes the storing of data in the form of cloud warehouses. Companies often use such a precise data mining method to have more in-depth real-time data analysis..

Few other applications can be:-

Applications	Usage
Communications	Data mining techniques are used in communication sector to predict customer behavior to offer highly targetted and relevant campaigns.
Insurance	Data mining helps insurance companies to price their products profitable and promote new offers to their new or existing customers.
Manufacturing	With the help of Data Mining Manufacturers can predict wear and tear of production assets. They can anticipate maintenance which helps them reduce them to minimize downtime.
Banking	Data mining helps finance sector to get a view of market risks and manage regulatory compliance. It helps banks to identify probable defaulters to decide whether to issue credit cards, loans, etc.
Retail	Data Mining techniques help retail malls and grocery stores identify and arrange most sellable items in the most attentive positions. It helps store owners to comes up with the offer which encourages customers to increase their spending.
Service Providers	Service providers like mobile phone and utility industries use Data Mining to predict the reasons when a customer leaves their company. They analyze billing details, customer service interactions, complaints made to the company to assign each customer a probability score and offers incentives.
E-Commerce	E-commerce websites use Data Mining to offer cross-sells and up-sells through their websites. One of the most famous names is Amazon, who use Data mining techniques to get more customers into their eCommerce store.
Super Markets	Data Mining allows supermarket's develope rules to predict if their shoppers were likely to be expecting. By evaluating their buying pattern, they could find woman customers who are most likely pregnant. They can start targeting products like baby powder, baby shop, diapers and so on.

Crime Investigation	Data Mining helps crime investigation agencies to deploy police workforce (where is a crime most likely to happen and when?), who to search at a border crossing etc.
Bioinformatics	Data Mining helps to mine biological data from massive datasets gathered in biology and medicine.

Benefits of Data Mining:

- Data mining technique helps companies to get knowledge-based information.
- Data mining helps organizations to make the profitable adjustments in operation and production.
- The data mining is a cost-effective and efficient solution compared to other statistical data applications.
- Data mining helps with the decision-making process.
- Facilitates automated prediction of trends and behaviors as well as automated discovery of hidden patterns.
- It can be implemented in new systems as well as existing platforms
- It is the speedy process which makes it easy for the users to analyze huge amount of data in less time.

Disadvantages of Data Mining

- There are chances of companies may sell useful information of their customers to other companies for money. For example, American Express has sold credit card purchases of their customers to the other companies.
- Many data mining analytics software is difficult to operate and requires advance training to work on.
- Different data mining tools work in different manners due to different algorithms employed in their design. Therefore, the selection of correct data mining tool is a very difficult task.
- The data mining techniques are not accurate, and so it can cause serious consequences in certain conditions.

DATA MINING PROJECTS

Today, data mining has become strategically important to organizations across industries. It not only helps in predicting outcomes and trends but also in removing bottlenecks and improving existing processes. But before we begin, let us look at an example to decode what data mining is all about. Suppose you have a data set containing login logs of a web application. It can include

things like the username, login timestamp, activities performed, time spent on the site before logging out, etc.

Such unstructured data in itself would not serve any purpose unless it is organized systematically and analyzed to extract relevant information for the business. By applying the different techniques of data mining, you can discover user habits, preferences, peak usage timings, etc. These insights can further increase the software system's efficiency and boost its user-friendliness. In today's digital era, the computing processes of collecting, cleaning, analyzing, and interpreting data make up an integral part of business strategies. So, data scientists are required to have adequate knowledge of methods like pattern tracking, classification, cluster analysis, prediction, neural networks, etc. The more you experiment with different data mining projects, the more knowledge you gain.

Few Data Mining Project Ideas :

1. GERF: Group Event Recommendation Framework

This is one of the simple data mining projects yet an exciting one. It is an intelligent solution for recommending social events, such as exhibitions, book launches, concerts, etc. A majority of the research focuses on suggesting upcoming attractions to individuals. So, a Group Event Recommendation Framework (GERF) was developed to propose events to a group of users.

This model uses a learning-to-rank algorithm to extract group preferences and can incorporate additional contextual influences with ease, accuracy, and time-efficiency. Also, it can be conveniently applied to other group recommendation scenarios like location-based travel services.

2. Protecting user data in profile-matching social networks

This is one of the convenient data mining projects that has a lot of use in the future. Consider the user profile database maintained by the providers of social networking services, such as online dating sites. The querying users specify certain criteria based on which their profiles are matched with that of other users. This process has to be secure enough to protect against any kind of data breaches. There are some solutions in the market today that use homomorphic encryption and multiple servers for matching user profiles to preserve user privacy.

3. Sentimental analysis and opinion mining for mobile networks

This project concerns post-publishing applications where a registered user can share text posts or images and also leave comments on posts. Under the prevailing system, users have to go through all the comments manually to filter out verified comments, positive comments, negative remarks, and so on.

With the sentiment analysis and opinion mining system, users can check the status of their post without dedicating much time and effort. It provides an opinion on the comments made on a post and also gives the option to view a graph.

4. Predicting consumption patterns with a mixture approach

Individuals consume a large selection of items in the digital world today. For example, while making purchases online, listening to music, using online navigation, or exploring virtual environments. Applications in these contexts employ predictive modeling techniques to recommend new items to users. However, in many situations, we want to know the additional details of previously-consumed items and past user behavior. And this is where the baseline approach of matrix factorization-based prediction falls short. This is one of the creative data mining projects.

A mixture model with repeated and novel events offers a suitable alternative for such problems. It aims to deliver accurate consumption predictions by balancing individual preferences in terms of exploration and exploitation. Also, it is one of those data mining project topics that include an experimental analysis using real-world datasets. The study's results show that the new approach works efficiently across different settings, from social media and music listening to location-based data.

TEXT MINING

Text Mining is one of the most critical ways of analyzing and processing unstructured data which forms nearly 80% of the world's data. Today a majority of organizations and institutions gather and store massive amounts of data in data warehouses, and cloud platforms and this data continues to grow exponentially by the minute as new data comes pouring in from multiple sources. As a result, it becomes a challenge for companies and organizations to store, process, and analyze vast amounts of textual data with traditional tools. This is where text mining applications, text mining tools, and text mining techniques come in.

Understanding TEXT MINING

According to *Wikipedia*, “Text mining, also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text.” The definition strikes at the primary chord of text mining – to delve into unstructured data to extract meaningful patterns and insights required for exploring textual data sources.

Text mining incorporates and integrates the tools of information retrieval, data mining, machine learning, statistics, and computational linguistics, and hence, it is nothing short of a multidisciplinary field. Text mining deals with natural language texts either stored in semi-structured or unstructured formats.

For businesses, the large amount of data generated every day represents both an opportunity and a challenge. On the one side, data helps companies get smart insights on people's opinions about a product or service. Think about all the potential ideas that you could get from analyzing emails, product reviews, social media posts, customer feedback, support tickets, etc. On the other side, there's the dilemma of how to process all this data. And that's where text mining plays a major

role. Like most things related to Natural Language Processing (NLP), text mining may sound like a hard-to-grasp concept. But the truth is, it doesn't need to be.

Difference between Text Mining, Text Analysis, and Text Analytics?

Text mining and text analysis are often used as synonyms. Text analytics, however, is a slightly different concept.

So, what's the difference between text mining and text analytics?

In short, they both intend to solve the same problem (automatically analyzing raw text data) by using different techniques. Text mining identifies relevant information within a text and therefore, provides qualitative results. Text analytics, however, focuses on finding patterns and trends across large sets of data, resulting in more quantitative results. Text analytics is usually used to create graphs, tables and other sorts of visual reports.

Text mining combines notions of statistics, linguistics, and machine learning to create models that learn from training data and can predict results on new information based on their previous experience.

Text analytics, on the other hand, uses results from analyses performed by text mining models, to create graphs and all kinds of data visualizations.

Choosing the right approach depends on what type of information is available. In most cases, both approaches are combined for each analysis, leading to more compelling results.

Methods and Techniques

There are different methods and techniques for text mining. In this section, we'll cover some of the most frequent.

Basic Methods

Word frequency

Word frequency can be used to identify the most recurrent terms or concepts in a set of data. Finding out the most mentioned words in unstructured text can be particularly useful when analyzing customer reviews, social media conversations or customer feedback.

For instance, if the words expensive, overpriced and overrated frequently appear on your customer reviews, it may indicate you need to adjust your prices (or your target market!).

Collocation

Collocation refers to a sequence of words that commonly appear near each other. The most common types of collocations are bigrams (a pair of words that are likely to go together, like get started, save time or decision making) and trigrams (a combination of three words, like within walking distance or keep in touch).

Identifying collocations — and counting them as one single word — improves the granularity of the text, allows a better understanding of its semantic structure and, in the end, leads to more accurate text mining results.

Concordance

Concordance is used to recognize the particular context or instance in which a word or set of words appears. We all know that the human language can be ambiguous: the same word can be used in many different contexts. Analyzing the concordance of a word can help understand its exact meaning based on context.

For example, here are a few sentences extracted from a set of reviews including the word ‘work’:

Concordance Example

Preceding context	Target	Following context
It saves time and helps teams	work	more efficiently.
Some advanced features only	work	in one language (English)
It enables us to	work	towards better conversion and retention.
We recommend this to several of the small businesses we	work	with, and they are all happy with the results.

Advanced Methods

Text Classification

Text classification is the process of assigning categories (tags) to unstructured text data. This essential task of Natural Language Processing (NLP) makes it easy to organize and structure complex text, turning it into meaningful data.

Thanks to text classification, businesses can analyze all sorts of information, from emails to support tickets, and obtain valuable insights in a fast and cost-effective way.

Below, we’ll refer to some of the most popular tasks of text classification – topic analysis, sentiment analysis, language detection, and intent detection.

- **Topic Analysis:** helps you understand the main themes or subjects of a text, and is one of the main ways of organizing text data. For example, a support ticket saying my online order hasn’t arrived, can be classified as Shipping Issues.

- **Sentiment Analysis:** consists of analyzing the emotions that underlie any given text. Suppose you are analyzing a series of reviews about your mobile app. You may find out that the most frequently mentioned topics in those reviews are UI-UX or Ease of Use, but that's not enough information to arrive to any conclusions. Sentiment analysis helps you understand the opinion and feelings in a text, and classify them as positive, negative or neutral. Sentiment analysis has a lot of useful applications in business, from analyzing social media posts to going through reviews or support tickets. In terms of customer support, for instance, you might be able to quickly identify angry customers and prioritize their problems first.
- **Language Detection:** allows you to classify a text based on its language. One of its most useful applications is automatically routing support tickets to the right geographically located team. Automating this task is quite simple and helps teams save valuable time.
- **Intent Detection:** you could use a text classifier to recognize the intentions or the purpose behind a text automatically. This can be particularly useful when analyzing customer conversations. For example, you could sift through different outbound sales email responses and identify the prospects which are interested in your product from the ones that are not, or the ones who want to unsubscribe.

Text Extraction

Text extraction is a text analysis technique that extracts specific pieces of data from a text, like keywords, entity names, addresses, emails, etc. By using text extraction, companies can avoid all the hassle of sorting through their data manually to pull out key information.

Most times, it can be useful to combine text extraction with text classification in the same analysis.

Below, we'll refer to some of the main tasks of text extraction – keyword extraction, named entity recognition and feature extraction.

- **Keyword Extraction:** keywords are the most relevant terms within a text and can be used to summarize its content. Utilizing a keyword extractor allows you to index data to be searched, summarize the content of a text or create tag clouds, among other things.
- **Named Entity Recognition:** allows you to identify and extract the names of companies, organizations or persons from a text.

- **Feature Extraction:** helps identify specific characteristics of a product or service in a set of data. For example, if you are analyzing product descriptions, you could easily extract features like color, brand, model, etc.

Why is Text Mining Important?

Individuals and organizations generate tons of data every day. Stats claim that almost 80% of the existing text data is unstructured, meaning it's not organized in a predefined way, it's not searchable, and it's almost impossible to manage. In other words, it's just not useful.

Being able to organize, categorize and capture relevant information from raw data is a major concern and challenge for companies. Text mining is crucial to this mission.

In a business context, unstructured text data can include emails, social media posts, chats, support tickets, surveys, etc. Sorting through all these types of information manually often results in failure. Not only because it's time-consuming and expensive, but also because it's inaccurate and impossible to scale.

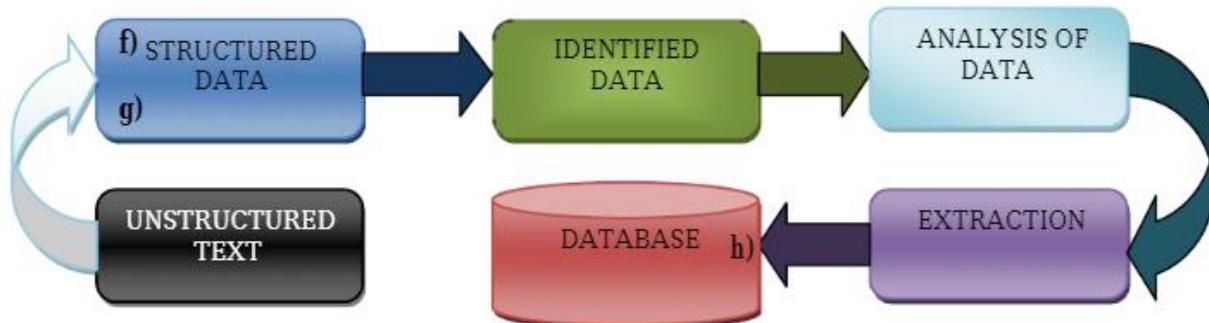
Text mining, however, has proved to be a reliable and cost-effective way to achieve accuracy, scalability and quick response times. Here are some of its main advantages in more detail:

- **Scalability:** with text mining it's possible to analyze large volumes of data in just seconds. By automating specific tasks, companies can save a lot of time that can be used to focus on other tasks. This results in more productive businesses.
- **Real-time analysis:** thanks to text mining, companies can prioritize urgent matters accordingly including, detecting a potential crisis, and discovering product flaws or negative reviews in real time. Why is this so important? Because it allows companies to take quick action.
- **Consistent Criteria:** when working on repetitive, manual tasks people are more likely to make mistakes. They also find it hard to maintain consistency and analyze data subjectively. Let's take tagging, for example. For most teams, adding categories to emails or support tickets is a time-consuming task that often leads to errors and inconsistencies. Automating this task not only saves precious time but also allows more accurate results and assures that a uniform criterion is applied to every ticket.

The five fundamental steps involved in text mining are:

1. Gathering unstructured data from multiple data sources like plain text, web pages, pdf files, emails, and blogs, to name a few.

2. Detect and remove anomalies from data by conducting pre-processing and cleansing operations. Data cleansing allows you to extract and retain the valuable information hidden within the data and to help identify the roots of specific words.
- For this, you get a number of text mining tools and text mining applications.
3. Convert all the relevant information extracted from unstructured data into structured formats.
 4. Analyze the patterns within the data via the Management Information System (MIS).
 5. Store all the valuable information into a secure database to drive trend analysis and enhance the decision-making process of the organization.



Text Mining Techniques

Text mining techniques can be understood at the processes that go into mining the text and discovering insights from it. These text mining techniques generally employ different text mining tools and applications for their execution. Now, let us now look at the various text mining techniques:

1. Information Extraction

This is the most famous text mining technique. Information exchange refers to the process of extracting meaningful information from vast chunks of textual data. This text mining technique focuses on identifying the extraction of entities, attributes, and their relationships from semi-structured or unstructured texts. Whatever information is extracted is then stored in a database for future access and retrieval. The efficacy and relevancy of the outcomes are checked and evaluated using precision and recall processes.

2. Information Retrieval

Information Retrieval (IR) refers to the process of extracting relevant and associated patterns based on a specific set of words or phrases. In this text mining technique, IR systems make use of different algorithms to track and monitor user behaviors and discover relevant data accordingly. Google and Yahoo search engines are the two most renowned IR systems.

3. Categorization

This is one of those text mining techniques that is a form of “supervised” learning wherein normal language texts are assigned to a predefined set of topics depending upon their content. Thus, categorization or rather Natural Language Processing (NLP) is a process of gathering text documents and processing and analyzing them to uncover the right topics or indexes for each document. The co-referencing method is commonly used as a part of NLP to extract relevant synonyms and abbreviations from textual data. Today, NLP has become an automated process used in a host of contexts ranging from personalized commercials delivery to spam filtering and categorizing web pages under hierarchical definitions, and much more.

4. Clustering

Clustering is one of the most crucial text mining techniques. It seeks to identify intrinsic structures in textual information and organize them into relevant subgroups or ‘clusters’ for further analysis. A significant challenge in the clustering process is to form meaningful clusters from the unlabeled textual data without having any prior information on them. Cluster analysis is a standard text mining tool that assists in data distribution or acts as a pre-processing step for other text mining algorithms running on detected clusters.

5. Summarisation

Text summarisation refers to the process of automatically generating a compressed version of a specific text that holds valuable information for the end-user. The aim of this text mining technique is to browse through multiple text sources to craft summaries of texts containing a considerable proportion of information in a concise format, keeping the overall meaning and intent of the original documents essentially the same. Text summarisation integrates and combines the various methods that employ text categorization like decision trees, neural networks, regression models, and swarm intelligence.

Technique	Characteristics	Tools
Retrieval	Retrieves valuable information from unstructured text	Intelligent Miner, Text Analyst
Extraction	Extract information from structured database	Text Finder, Clear Forest Text
Summarization	Reduce length by keeping its main points and overall meaning as it is	Tropic Tracking Tool, Sentence Ext Tool
Categorization	Document based categorization	Intelligent Miner
Cluster	Cluster collection of documents, Clustering, classification and analysis of text document	Carrot, Rapid Miner

APPLICATIONS OF TEXT MINING

Text mining techniques and text mining tools are rapidly penetrating the industry, right from academia and healthcare to businesses and social media platforms. This is giving rise to a number of text mining applications. Here are a few text mining applications used across the globe today:

1. Risk Management

One of the primary causes of failure in the business sector is the lack of proper or insufficient risk analysis. Adopting and integrating risk management software powered by text mining technologies such as SAS Text Miner can help businesses to stay updated with all the current trends in the business market and boost their abilities to mitigate potential risks. Since text mining tools and technologies can gather relevant information from across thousands of text data sources and create links between the extracted insights, it allows companies to access the right information at the right moment, thereby enhancing the entire risk management process.

2. Customer Care Service

Text mining techniques, particularly NLP, are finding increasing importance in the field of customer care. Companies are investing in text analytics software to enhance their overall customer experience by accessing the textual data from varied sources such as surveys, customer feedback, and customer calls, etc. Text analysis aims to reduce the response time of the company and help address the grievances of the customers speedily and efficiently.

3. Fraud Detection

Text analytics backed by text mining techniques provides a tremendous opportunity for domains that gather a majority of data in the text format. Insurance and finance companies are harnessing this opportunity. By combining the outcomes of text analyses with relevant structured data these companies are now able to process claims swiftly as well as to detect and prevent frauds.

4. Business Intelligence

Organizations and business firms have started to leverage text mining techniques as part of their business intelligence. Apart from providing profound insights into customer behavior and trends, text mining techniques also help companies to analyze the strengths and weaknesses of their rivals, thus, giving them a competitive advantage in the market. Text mining tools such as Cogito Intelligence Platform and IBM text analytics provide insights on the performance of marketing strategies, latest customer and market trends, and so on.

5. Social Media Analysis

There are many text mining tools designed exclusively for analyzing the performance of social media platforms. These help to track and interpret the texts generated online from the news, blogs, emails, etc. Furthermore, text mining tools can efficiently analyze the number of posts, likes, and followers of your brand on social media, thereby allowing you to understand the reaction of people who are interacting with your brand and online content. The analysis will enable you to understand ‘what’s hot and what’s not’ for your target audience.

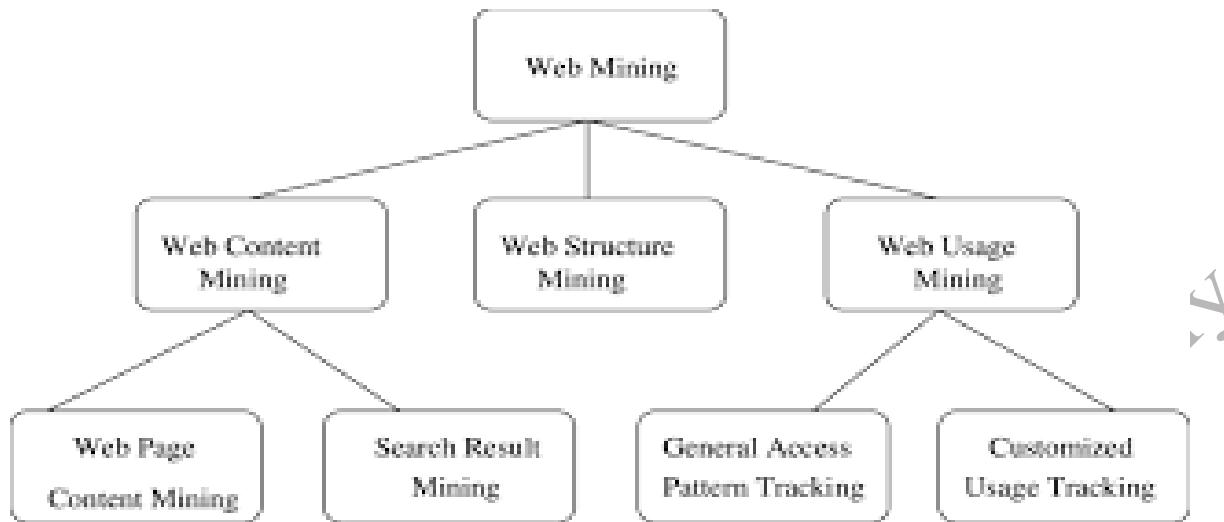
WEB MINING

Web Mining is the process of Data Mining techniques to automatically discover and extract information from Web documents and services. The main purpose of web mining is discovering useful information from the World-Wide Web and its usage patterns. Web mining aims to discover and retrieve useful and interesting patterns from large data sets, as well as in the classic data mining. Big data act as data sets on web mining. Web data includes information, documents, structure and profile. Web mining is based on two concepts defined, process-based and data-driven. (Based on data routinely and commonly used).

In the view of Web mining data web is used to extract knowledge . In general, the use of web mining typically involves several steps: collecting data, selecting the data before processing, knowledge discovery and analysis.

Types of web mining

Web mining can be generally divided into three categories, as seen below:



- **Web content mining**

Web content mining is the process of extracting useful information from the content of Web documents. The contents of a web document is corresponding to the concepts that the document sought to transfer it to users. This content can include text, image, video, sound or records such as lists and tables. The text mining has been studied more than other areas.

- **Web structure mining**

The web can be represented as graph which its nodes and edges are the links between documents. Web structure mining is the process of extracting structural information from the web.

- **Web usage mining**

Web usage mining is the application of data mining techniques to discover patterns using the Web to better understand and meet the needs of the user. This type of web mining explores data relating to the use of web users. It should be noted that there are no clear boundaries between web mining groups. For example, web content mining techniques can use user information in addition to using the documents. It can also be achieved to better results by the combination of above techniques

APPLICATIONS OF WEB MINING:

1. Web mining helps to improve the power of web search engine by classifying the web documents and identifying the web pages.
2. It is used for Web Searching e.g., Google, Yahoo etc and Vertical Searching e.g., FatLens, Become etc.
3. Web mining is used to predict user behavior.
4. Web mining is very useful of a particular Website and e-service e.g., landing page optimization.

Web mining can be broadly divided into three different types of techniques of mining: Web Content Mining, Web Structure Mining, and Web Usage Mining. These are explained as following below.

Web Content Mining:

1. **Web content mining** is the application of extracting useful information from the content of the web documents. Web content consist of several types of data – text, image, audio, video etc. Content data is the group of facts that a web page is designed. It can provide effective and interesting patterns about user needs. Text documents are related to text mining, machine learning and natural language processing. This mining is also known as text mining. This type of mining performs scanning and mining of the text, images and groups of web pages according to the content of the input.

2. Web Structure Mining:

Web structure mining is the application of discovering structure information from the web. The structure of the web graph consists of web pages as nodes, and hyperlinks as edges connecting related pages. Structure mining basically shows the structured summary of a particular website. It identifies relationship between web pages linked by information or direct link connection. To determine the connection between two commercial websites, Web structure mining can be very useful.

3. Web Usage Mining:

Web usage mining is the application of identifying or discovering interesting usage patterns from large data sets. And these patterns enable you to understand the user behaviors or something like that. In web usage mining, user access data on the web and collect data in form of logs. So, Web usage mining is also called log mining.

Comparison Between Data mining and Web mining:

Points	Data Mining	Web Mining
Definition	Data Mining is the process that attempts to discover pattern and hidden knowledge in large data sets in any system.	Web Mining is the process of data mining techniques to automatically discover and extract information from web documents.
Application	Data Mining is very useful for web page analysis.	Web Mining is very useful for a particular website and e-service.
Target Users	Data scientist and data engineers.	Data scientists along with data analysts.

Points	Data Mining	Web Mining
Access	Data Mining is access data privately.	Web Mining is access data publicly.
Structure	In Data Mining get the information from explicit structure.	In Web Mining get the information from structured, unstructured and semi-structured web pages.
Problem Type	Clustering, classification, regression, prediction, optimization and control.	Web content mining, Web structure mining.
Tools	It includes tools like machine learning algorithms.	Special tools for web mining are Scrapy, PageRank and Apache logs.
Skills	It includes approaches for data cleansing, machine learning algorithms. Statistics and probability.	It includes application level knowledge, data engineering with mathematical modules like statistics and probability.

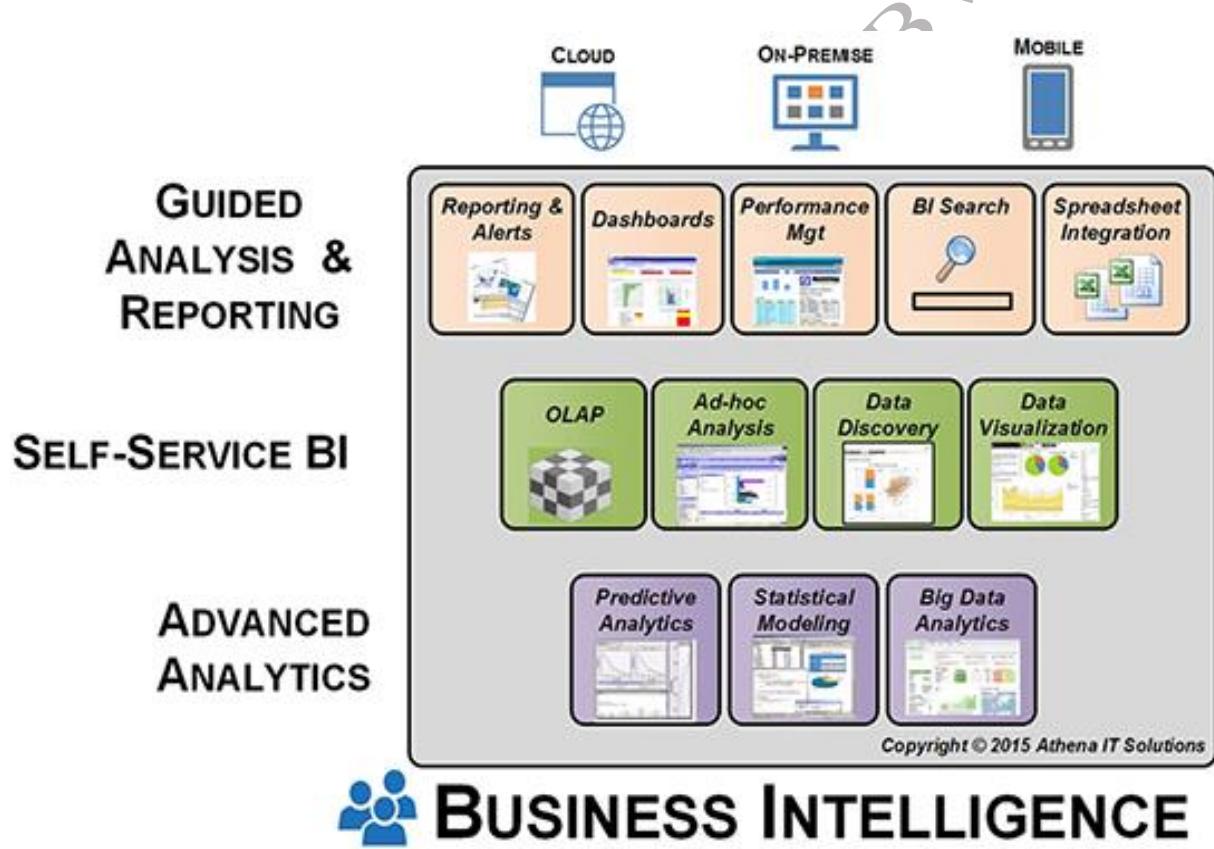
For Internal Circulation Only

UNIT IV: APPLICATIONS OF BUSINESS ANALYTICS

Business Analytics

What's the difference between Business Analytics and Business Intelligence?

The first is the business aspect of BI — the need to get the most value out of information. This need hasn't really changed in over fifty years (although the increasing complexity of the world economy means it's ever harder to deliver). And the majority of real issues that stop us from getting value out of information (information culture, politics, lack of analytic competence, etc.) haven't changed in decades either.



The second is the IT aspect of BI — what technology is used to help provide the business need. This obviously does change over time — sometimes radically.

The problems in nomenclature typically arise because “business intelligence” is commonly used to refer both of these, according to the context, thus confusing the heck out of everyone.

In particular, as the IT infrastructure inevitably changes over time, analysts and vendors (especially new entrants) become uncomfortable with what increasingly strikes them as a “dated” term, and want to change it for a newer term that they think will differentiate their coverage/products (when I joined the industry, it was called “decision support systems” – which I still think is a better term in many ways).

When people introduce a new term, they inevitably (and deliberately, cynically?) dismiss the old one as “just technology driven” and “backward looking”, while the new term is “business oriented” and “actionable”.

➤ **Discriminant Analysis:**

Discriminant Analysis may be used for two objectives: either we want to assess the adequacy of classification, given the group memberships of the objects under study; or we wish to assign objects to one of a number of (known) groups of objects. Discriminant Analysis may thus have a descriptive or a predictive objective.

In both cases, some group assignments must be known before carrying out the Discriminant Analysis. Such group assignments, or labelling, may be arrived at in any way. Hence discriminant Analysis can be employed as a useful complement to Cluster Analysis (in order to judge the results of the latter) or Principal Components Analysis.

The major underlying assumptions of DA are:

- the observations are a random sample;
- each predictor variable is normally distributed;
- each of the allocations for the dependent categories in the initial classification are correctly classified;
- there must be at least two groups or categories, with each case belonging to only one group so that the groups are mutually exclusive and collectively exhaustive (all cases can be placed in a group)

Application - Understanding and modeling differences between / among groups (e.g., buyers vs. non-buyers of different brands); Predicting market behavior based on demographic and psychographic variables.

➤ **Cluster Analysis:**

Cluster analysis is an exploratory data analysis tool for solving classification problems. Its object is to sort cases (people, things, events, etc.) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. A cluster is a group of relatively homogenous cases or observations. Each cluster thus describes, in terms of the data collected, the class to which its members belong; and this description may be abstracted through use from the particular to the general class or type. Uses any of several techniques (viz. Nearest Neighbors, K-Means etc.) to classify people, objects, or variables into more homogeneous groups.

Application - Identifying / describing market segments; developing typological findings and describing target markets.

Pros - Allows a deeper understanding of the market. Can greatly aid messaging and new product development by targeting homogeneous groups.

Cons - Subjective interpretation of the results is a component. The technique is mathematical and therefore has no underlying model against which to test statistical hypotheses. K-means is a fast cluster analysis method, in which accuracy depends on the use of initialization algorithms that are usually serial and slow.

Predictive Analytics

Organizations are turning to predictive analytics to increase their bottom line and competitive advantage. Some of the most common uses include:

Detecting fraud. Combining multiple analytics methods can improve pattern detection and prevent criminal behavior. As cybersecurity becomes a growing concern, high-performance behavioral analytics examines all actions on a network in real time to spot abnormalities that may indicate fraud, zero-day vulnerabilities and advanced persistent threats.

Optimizing marketing campaigns. Predictive analytics are used to determine customer responses or purchases, as well as promote cross-sell opportunities. Predictive models help businesses attract, retain and grow their most profitable customers.

Improving operations. Many companies use predictive models to forecast inventory and manage resources. Airlines use predictive analytics to set ticket prices. Hotels try to predict the number of guests for any given night to maximize occupancy and increase revenue. Predictive analytics enables organizations to function more efficiently.

Reducing risk. Credit scores are used to assess a buyer's likelihood of default for purchases and are a well-known example of predictive analytics. A credit score is a number generated by a

predictive model that incorporates all data relevant to a person's creditworthiness. Other risk-related uses include insurance claims and collections.

➤ Logistic Regression

Regression analysis is a form of predictive modeling technique which investigates the relationship between a dependent (target) and independent variable (s) (predictor). This technique is used for forecasting, time series modeling and finding the causal effect relationship between the variables. For example, relationship between rash driving and number of road accidents by a driver is best studied through regression.

Regression analysis is an important tool for modeling and analyzing data. Here, we fit a curve / line to the data points, in such a manner that the differences between the distances of data points from the curve or line is minimized.

Application – regression analysis estimates the relationship between two or more variables. Let's understand this with an easy example:

Let's say, you want to estimate growth in sales of a company based on current economic conditions. You have the recent company data which indicates that the growth in sales is around two and a half times the growth in the economy. Using this insight, we can predict future sales of the company based on current & past information.

There are multiple benefits of using regression analysis. They are as follows:

- It indicates the significant relationships between dependent variable and independent variable.
- It indicates the strength of impact of multiple independent variables on a dependent variable.

Regression analysis also allows us to compare the effects of variables measured on different scales, such as the effect of price changes and the number of promotional activities. These benefits help market researchers / data analysts / data scientists to eliminate and evaluate the best set of variables to be used for building predictive models.

Logistic Regression

It can be argued that the most important step in a business analytics process is establishing a clear business objective. Once this is done, selecting the right technique becomes a matter of simple logic. At a very high level there are fundamentally two main classes of techniques: those that evolved purely from statistics (such as regression) and those that emerged from a blend of stats, computer science and mathematics (such as classification trees).

Logistic regression is used to find the probability of event=Success and event=Failure. We should use logistic regression when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature. The key assumptions here are that both the predictor and target variables are continuous as seen in this chart below. Intuitively, one can state that when X increases, Y increases along the slope of the line.

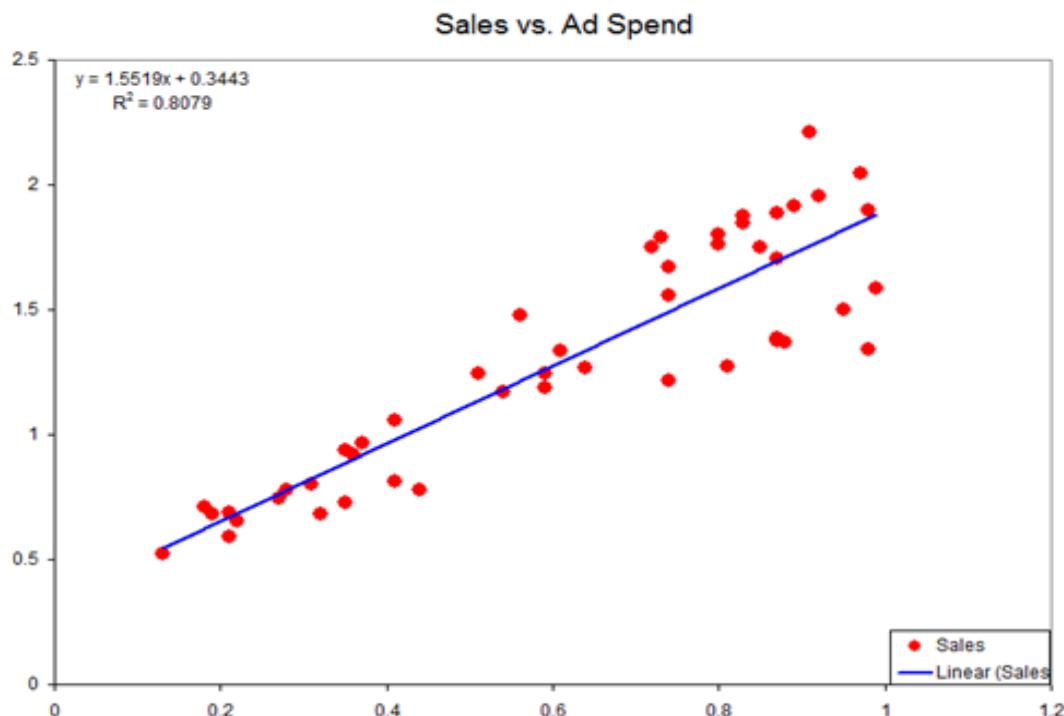


Fig 1. Linear Regression Model. We can make an intuitive assessment that increase in *Ad spend* also increases *Sales*. Using the straight line, we may also be able to predict.

What happens if the target variable is not continuous? When the target (Y) variable is discrete, the straight line is no longer a fit as seen in this chart. Although intuitively we can still state that when X (say advertising spend) increases, Y (say response or no response to a mailing

campaign) also increases, but there is no gradual transition, the Y value abruptly jumps from one binary outcome to the other. Thus the straight line is a poor fit for this data.

Important Points:

- It is widely used for classification problems
- Logistic regression doesn't require linear relationship between dependent and independent variables. It can handle various types of relationships because it applies a non-linear log transformation to the predicted odds ratio
- To avoid over fitting and under fitting, we should include all significant variables. A good approach to ensure this practice is to use a step wise method to estimate the logistic regression
- It requires large sample sizes because maximum likelihood estimates are less powerful at low sample sizes than ordinary least square
- The independent variables should not be correlated with each other i.e. no multi collinearity. However, we have the options to include interaction effects of categorical variables in the analysis and in the model.
- If the values of dependent variable is ordinal, then it is called as Ordinal logistic regression
- If dependent variable is multi class then it is known as Multinomial Logistic regression.

➤ **Principal Component Analysis**

Principal component analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. It's often used to make data easy to explore and visualize. The goal of principal components analysis is to explain the maximum amount of variance with the fewest number of principal components. Principal components analysis is commonly used in the social sciences, market research, and other industries that use large data sets.

Principal components analysis is commonly used as one step in a series of analyses. You can use principal components analysis to reduce the number of variables and avoid

multicollinearity, or when you have too many predictors relative to the number of observations.

Example

A consumer products company wants to analyze customer responses to several characteristics of a new shampoo: color, smell, texture, cleanliness, shine, volume, amount needed to lather, and price. They perform a principal components analysis to determine whether they can form a smaller number of uncorrelated variables that are easier to interpret and analyze. The results identify the following patterns:

- Color, smell, and texture form a "Shampoo quality" component.
- Cleanliness, shine, and volume form an "Effect on hair" component.
- Amount needed to lather and price form a "Value" component.

RISK AND FRAUD DETECTION AND PREDICTION IN INDIA

Fraud and abuse are constant challenges for online businesses. The digital explosion has brought many complex and cross channel risks to business than ever before. Typically, fraudsters attempt to get past various cross-channel and security protocols with stolen data and credentials obtained through sim swaps, man-in-the-middle attacks, phishing keylogging or password guessing attacks. They can also buy sensitive data from the darknet with your data housed in separate silos or residing with the third-party digital platforms. Apart from that, frauds happen with credit-based financial businesses with deliberate loan delinquency and similar malicious behaviours from some users.

Such instances can create severe financial and reputational damages for companies and, therefore, it is important to get real-time fraud analytics across digital channels. India being one of the largest digital economies in the world, there are many firms launching analytics-based fraud solutions in the market for not just financial companies but across different industries.

Fraud detection is a challenging problem. The fact is that fraudulent transactions are rare; they represent a very small fraction of activity within an organization. The challenge is that a small percentage of activity can quickly turn into big dollar losses without the right tools and systems in place. Criminals are crafty. As traditional fraud schemes fail to pay off, fraudsters have learned to change their tactics. The good news is that with advances in fraud analytics, systems can learn, adapt and uncover emerging patterns for preventing fraud.

Indian Banking sector detected frauds worth more than the Rs 71,000 crore recapitalisation package planned by the government.

The payments industry is evolving, and so are consumer spending habits and expectations. On one hand POS credit transactions grew by 29.8% y-o-y, while debit card swipes increased by 14.4%, taking market share from cash payments. On the other hand, UPI transactions continued to witness increasing merchant acceptance and crossed 100 million users, allowing the e-payments ecosystem to thrive.

In that sense, 2019 has been yet another significant year for the payments industry in India and the pace of evolution here has gone into overdrive. And it is expected to only get faster. This makes it critical for payments companies, banks and other players in the space to anticipate what's next—now.

Unfortunately, fraudsters have also been keeping pace with these developments, constantly in search of weaknesses to exploit. The latest RBI report quoted that the banking system in India detected frauds worth Rs 71,500 crore in the financial year 2018-19.

To put this into context, this is slightly more than the Rs 71,000 crore recapitalization package planned by the government to revive the health of the public sector banks! Hence, it is essential for banks – and all the parties involved in the payments ecosystem – to constantly adapt to emerging fraud threats.

For example, if a user logs into his or her account at an irregular rate or suddenly begins adding priority shipping to high-priced orders, the system will detect this irregularity and immediately raise a red flag. However, if a user simply purchases an expensive gift, or books travel arrangements during the festive season—behaviours that coincide with seasonal activity—the system will recognize and differentiate the fraudulent from the legitimate accordingly.

Democratized machine learning: With ‘democratized’ machine learning, what we mean is that financial crime risk managers, with specific subject matter expertise, are equipped with machine learning modelling tools that they can build, test and deploy on their own – without extensive help from tech experts. This will give them better understanding and more power to explain their choice of features, use of model scores and the actions taken to their management, auditors and regulators. This more democratized approach to machine learning will make it easier to not only address specific threats but to show ROI – one of the biggest reasons why this trend will emerge strongly.

Centralized payments intelligence hub: With real-time payments, the window for fraud prevention is much shorter and the ability to recover a fraudulent payment is much lower. Traditional rules-only systems are great at detecting known threats but can't uncover new criminal fraud strategies or zero-day attacks, which puts customers at risk. Hence, 2020 will see more interest among banks for shared intelligence (features and signals vs. data) for increased accuracy of fraud detection and prevention. In fact, The Reserve Bank of India (RBI), in August this year proposed to set up a central payment fraud registry with an intention to monitor digital payment frauds in real-time – a move that has been lauded by the industry. A common repository will definitely help banks work together rather than in silos when it comes to dealing with frauds, especially where the money is siphoned off across various bank accounts in different geographies.

Furthermore, the trends and patterns from the repository will also be used to improve banks' analytics engines and help in predicting future frauds.

Rise of application fraud: Across the world, identity scans are largely broken, meaning that synthetic IDs and pure identity theft will continue to increase next year, especially as banks and credit grantors continue to neglect reporting these losses or lose them in credit losses. This could be further fuelled by attacks on central infrastructures that manage digital identities and other important information, for example attacks on Aadhaar data and similar breaches.

Card-not-present (CNP) fraud: CNP fraud is done by obtaining details of a credit card holder like billing address, account number, three-digit security code and expiry date of the card. Credit card holders are generally fooled through mediums like online phishing, but both customers and merchants suffer when card-not present fraud occurs. In the recent past, after the mandate of two-factor authentication from RBI, the number of such 'card-less' or 'card not present' transactions have marginally gone down. However, due to the rise in SIM swaps and skimming, the percentage of such frauds could potentially rocket. These types of attacks will compel businesses to look beyond traditional endpoint security solutions if they are going to effectively combat evolving threats.

Cybersecurity and digital payment fraud cases are a critical concern when it comes to digitization of payments, and banks and other players in the ecosystem will have to take the right steps to mitigate them. Investments in next-level authentication methods, behavioural biometrics, multi-factor authentication, and real-time monitoring of frauds are some of the tools that will come to the fore in 2020 and can be leveraged to reduce frauds and continue to instil in consumers' confidence in digital payments.

India poses high risks of corruption and money laundering for financial institutions operating in the country. To evaluate these risks, the authors recommend four key general steps combined with a number of India-specific considerations. Anti-corruption efforts, they suggest, require, among other things, a comprehensive knowledge and understanding of all customers and business affiliates, including third parties and joint venture partners. For anti-money laundering, the authors emphasize "Know Your Customer" procedures that meet the Reserve Bank of India's robust standards, along with a strong monitoring program to help detect common money laundering schemes. For years, multinational companies have flocked to India to take advantage of the country's low-cost labor force and its burgeoning middle and consumer classes. Recently, India's emerging markets have attracted even more attention with the election in May 2014 of business-friendly Prime Minister Narendra Modi, who has promised to boost infrastructure spending and job growth while combating high inflation rates and bureaucratic red tape. India's economy is expected to improve significantly over the next several years, with some experts forecasting that it could soon outpace China as the world's premiere growth market. Indeed, financial institutions should take note — the formalization of India's economy under Prime Minister Modi and the corresponding surge in foreign business investment may create lucrative opportunities for banks, lenders, and money services businesses alike. India remains a challenging country in which to conduct business. For example, corruption continues to be a vexing problem, while the informal economy renders financial institutions common targets for money laundering. Prime Minister

Modi has promised to combat corruption by significantly enhancing local law enforcement resources and related prosecutions in India—practices largely absent from the country’s prior enforcement regimes. In light of this pledge by India’s new leadership, and considering that financial institutions already face a heightened level of global enforcement with respect to both anti-corruption and anti-money laundering compliance—a trend likely to continue as new laws are passed and cooperation between international authorities increases—financial institutions must ensure that they have a robust, comprehensive compliance program firmly in place before committing to business in India. Risk assessment is an integral part of any strong anti-corruption and anti-money laundering compliance program. The practice of regularly evaluating risk not only allows financial institutions to detect potential violations of anti-corruption and anti-money laundering laws and regulations, it also plays a key role in helping companies shape their compliance programs to prevent future violations and related improprieties. Noting the considerable challenges of conducting business in India and the anticipated uptick in anti-corruption and anti-money laundering enforcement there, the purpose of this article is to spotlight best practices for conducting risk assessments in India in order to help financial institutions enhance their compliance programs and increase the prospect of satisfying the expectations of law enforcement and regulatory authorities in India and around the world.

OVERVIEW OF RISK IN INDIA

Financial institutions must be prepared to manage significant corruption and money laundering risks if they intend to take advantage of the expanding Indian business landscape. A thorough appraisal of these risks, as unique to each company’s specific business and operational profiles, is critical to conducting a meaningful risk assessment.

Corruption Risk

India has long suffered from a deeply entrenched culture of corruption. Graft remains a strong part of Indian culture today. The country’s vast and complex bureaucracy creates abundant opportunities for government officials to demand bribe payments in exchange for assistance or benefits. The Indian government employs a large number of government workers, about 90% of which are classified as “low-wage earners.” The government often entrusts these particular employees with the power to grant or deny many of the licenses and permits that companies need in order to operate in the country. Because these workers are paid so little, many rely on bribes to supplement their income. Several key touch points between companies and the government in India therefore become opportunities for corruption to proliferate. Indeed, the sheer number of licenses and permits required by the government for multinational corporations to operate in many business sectors, including the financial sector, ensures that such opportunities are plentiful. According to a recent survey conducted by Transparency International, a leading non-profit organization dedicated to stopping corruption and promoting accountability and integrity, 80% of Indians believe that corruption in the public sector is “a problem” or “a serious problem.” Only 3% of respondents said that corruption was “not really a problem” or “not a problem at all.”

Corruption also appears to be trending in the wrong direction, with 71% of respondents opining that the level of corruption in India increased over the past two years, compared to only 7% who observed a decrease. Unsurprisingly, India has become a hot spot for global anti-corruption enforcement actions, particularly those pursued by the U.S. government under the U.S. Foreign Corrupt Practices Act. The FCPA prohibits bribery of foreign (i.e., non-U.S.) government officials and requires publicly traded companies in the United States to maintain accurate books and records and effective internal controls. In the past 10 years, companies have collectively paid billions of dollars to settle FCPA matters with the U.S. government. And several of the higher-profile FCPA enforcement actions have involved India. For example, in 2011, Diageo paid over \$16 million to the U.S. Securities and Exchange Commission to settle a matter involving allegations that its Indian distributors paid bribes to employees of government-owned liquor stores to increase beverage purchases. In 2010, Pride International paid \$56.2 million to the SEC and the U.S. Department of Justice in response to allegations that it paid bribes to an Indian judge to secure a favorable determination in a customs dispute.

Money Laundering Risk

India's extensive informal economy poses a significant money laundering risk to financial institutions operating in the country. Many wealthy Indians go to significant lengths to keep their assets out of regulated transactions or savings accounts. Some of these individuals may have earned their assets through criminal activity, while others simply do not want to pay taxes on their income. To avoid the formal economy, wealth is stored and transferred in alternate asset classes, such as land, housing, or precious metals, which can be difficult to trace for law enforcement purposes. Furthermore, approximately 85% of employed Indians are paid in cash. A staggering amount of India's economy exists outside of the regulated financial system — Indian regulators recently declared that the so-called "black economy" is estimated to be worth as much as \$2 trillion, more than India's annual gross domestic product. Some prosperous Indians have also turned to the offshore finance system to disguise the source and amount of their overall wealth. Experts have estimated that anywhere from \$500 billion to \$1.4 trillion in illicit Indian money has flowed out of the country into foreign bank accounts since 1948. Some Indians send "black money" out of the country only to arrange to have it transferred back into local financial institutions disguised as legitimate foreign funds. Until recently, the Indian government did little to curtail this practice. Anti-money laundering enforcement, however, is now on the rise. In June 2010, for example, India was admitted as the 34th country- member of the Financial Action Task Force ("FATF"), an inter-governmental body that monitors financial crime in its member countries and recommends government action to combat money laundering, terrorist financing, and other threats to the integrity of the international financial system. The Reserve Bank of India ("RBI") has also taken greater responsibility in fighting money laundering. The RBI has issued several advisories reiterating for local financial institutions their anti-money laundering obligations, and it has fined many banks for failure to comply with relevant Indian regulations.

BEST PRACTICES FOR RISK ASSESSMENT IN INDIA

In the wake of the significant anti-corruption and anti-money laundering compliance challenges faced by financial institutions seeking to enter or expand in the Indian market, there are a number of best practices that are effective for purposes of assessing and mitigating risk. These best practices are informed by the standards reflected in the various regulatory regimes of the DOJ, the SEC, the World Bank, and the Organisation for Economic Co-operation and Development (“OECD”). Importantly, as with most compliance procedures, an appropriate risk assessment strategy must be tailored to the particular risk profile of the business conducted by the company and the company’s operational footprint. Broadly speaking, financial institutions should conduct periodic, informed, and documented risk assessments with an emphasis on the areas of highest risk within the company’s business and operational structure. Geography is a key factor in evaluating risk, and the conduct of financial operations in India unquestionably qualifies as an elevated risk consideration. Risk assessment generally requires four principal steps. The first step is for the financial institution to determine the scope of the review and gather relevant information about its business operations and practices. For operations in India, it is particularly important to concentrate at the outset on the company’s client base, top-level commitment, existing standards and controls, and training and response protocols. Another important component of this step is to identify all laws and regulations that apply to the company and the local authorities with jurisdiction over the company’s activities. This component is particularly germane in India, where the shifting political landscape has led to changes in the government’s regulatory regime, such as the creation of the Indian Supreme Court’s Special Investigation Team (“SIT”) in 2014. The second step is to conduct relevant interviews of key company officers, employees, and other stakeholders with strong knowledge of the company’s operations. Importantly, this would include personnel on the ground in India who are involved in the company’s day-to-day business activities there. To conduct effective, focused, and streamlined interviews, relevant documents and materials should be collected and reviewed in advance. The third step is to carefully evaluate information gathered from the document review and interview stages, and begin to develop a report that underscores the company’s risk profile and outlines recommendations formulated to enhance the company’s compliance program by addressing determined risks (sometimes called compliance “gaps”). During this third step, the information collected over the course of the earlier phases should be scrutinized to identify “red flags” within the company’s risk profile and pinpoint areas of risk that should be attended to on a priority basis. Note that issues or “gaps” requiring even further review or investigation may be found in step three, and, if so, supplemental evaluation should be conducted, as necessary and appropriate, during this phase of the risk assessment process. The fourth and concluding step of a risk assessment in India is to compile and synthesize the findings of the review in a final report, which must include a plan for designing and implementing the recommended anti-corruption and/or anti-money laundering program enhancements. The final report should contain very specific suggestions for improving the financial institution’s compliance program and prioritize those recommendations based on the nature of the risk and the complexity or difficulty of design and implementation. The report should conclude with a detailed action plan assigning responsibility for each enhancement to a specific department or individual (with corresponding timeframes for completion).

Anti-Corruption Risk Assessment Considerations

While the four general steps detailed above are designed to assist financial institutions with business operations in India and around the world to manage risk, there are certain additional factors particularly relevant to India that should be considered with respect to anti-corruption risk, on the one hand, and anti-money laundering risk, on the other. Combining the four general steps with consideration of India-specific factors will help financial institutions maximize the customization of the assessment process to the specific type of risk under review. For example, perhaps the most critical factor in conducting an effective anti-corruption risk assessment in India relates to identifying the level of business activity with government officials or state-owned or-operated entities. Indeed, under the FCPA, the U.K. Bribery Act, and other anti-bribery laws around the world (such as Brazil's recently enacted Clean Company Act), bribing foreign government officials can create substantial criminal and/or civil and administrative liability. Financial institutions and other companies operating in India, therefore, must have a comprehensive knowledge and understanding of all customers and business affiliates, including third parties and joint venture partners. For a company to maximize its understanding of government-related risk, the assessment should seek up front to determine if any customers or business partners are government officials of state-owned or -operated entities. If so, the financial institution will want to include a review of the effectiveness of current internal controls relating to the treatment and oversight of those relationships. This would incorporate, for instance, careful consideration of the company's compensation arrangements with government partners, including commissions and profit margins that each partner receives as part of its current contractual arrangement. Importantly, however, an anti-corruption risk assessment in India must identify all touch points between the financial institution and government officials — not just those relating to contractual payments. Given the substantial risk that low-wage officials in India may demand bribe payments in exchange for licenses, permits, or other basic services, the assessment should consider the safeguards surrounding all of the company's communications with government officials. The assessment should also seek to ensure that the company maintains records that carefully document contacts with the Indian government, and any fees or costs paid to procure related services, including expenses incurred involving other payments to government officials. Further, an anti-corruption assessment should examine the company's compliance culture with respect to its operations in India (and, ultimately, around the world). Authorities in the United States, the United Kingdom, and elsewhere expect local company officers and onsite managers stationed in India to set the tone for anti-corruption compliance, and to detect and address inappropriate or unethical practices before they become major compliance issues. Given the pervasiveness of corruption in India, it is particularly important during the interview stage of the assessment to determine whether the company has a strong compliance culture, beginning with a positive "tone at the top" emanating from officers and managers to employees across operations in India. A robust anti-corruption culture will incorporate regular anti-corruption training and easily accessible policies, procedures, and response and escalation protocols.

Anti-Money Laundering Risk Assessment Considerations

As with anti-corruption evaluations, risk assessments in the anti-money laundering realm also have important and distinctive components that ultimately shape their utility and effectiveness. Importantly, within the context of observing and adhering to the general best practice risk assessment steps set forth above, a specific anti-money laundering risk assessment for financial institutions (and other multinational companies) operating in India should be designed to review products, services, customers, entities, and locations unique to the company's risk profile and that may be particularly vulnerable to money laundering activity. Although attempts to launder money, finance terrorism, or conduct related illicit activities through a financial institution can originate from many disparate sources, certain products, services, customers, entities, and geographic locations, such as India, are more susceptible. These include, for example, electronic funds payment services, private banking, trust and asset management services, issuance of monetary instruments, foreign correspondent accounts, trade finance, foreign exchange activity, services provided to third-party payment processors or senders, and non-deposit account services (such as investment products or insurance). To facilitate the review of potentially problematic customers, products, entities, and services as part of a regular risk assessment process, it is imperative for financial institutions in India to ensure that the company has a comprehensive transaction monitoring system in place to track activity within all of these areas. Indeed, financial institutions operating in India must comply with certain reporting requirements, including the filing of Suspicious Transaction Reports ("STRs"). To fulfill such obligations, they need a vigorous monitoring structure that can identify and sift through large amounts of data to generate meaningful real-time alerts that can be swiftly evaluated and acted upon. A strong monitoring program will also help detect common money laundering schemes in India, such as transfers of illicit funds from offshore accounts. An assessment of money laundering risk in India must also be designed to determine whether the company is in compliance with the RBI's "Know Your Customer" ("KYC") guidelines. Similar to the anti-money laundering regulatory regimes of other countries, a key objective of the RBI's KYC guidelines is to prevent financial institutions from being used by criminal elements for money laundering or terrorist financing activities. Robust KYC procedures enable financial institutions in India to analyze and, ultimately, understand the profiles of their customers and related financial dealings. This, in turn, helps them sensibly manage risks. Financial institutions in India must be particularly careful to identify customers that are foreign financial institutions, senior political figures, foreign corporations, deposit brokers, cash-intensive businesses, and professional service providers (such as consultants). Among other things, the RBI's KYC guidelines require financial institutions to follow certain customer identification procedures designed "to obtain sufficient information necessary to establish, to their satisfaction, the identity of each new customer, whether regular or occasional, and the purpose of the intended nature of the banking relationship." This obligation includes, for example, determination of the beneficial owners of each account (the person on whose behalf the account is ultimately maintained), screening for Politically Exposed Persons ("PEPs"), and special attention for accounts opened by "professional intermediaries." Accordingly, an anti-money laundering risk assessment in India must ensure that the financial institution's current KYC program is designed to obtain sufficient data to verify the identity of each and every customer — including address and photograph — to be certain that the RBI's screening priorities are satisfied. Those conducting the

assessment should seek to confirm that, for all customers, the company verifies the legal status of the person or entity through proper documents, whether any person purporting to act on behalf of the legal person or entity is authorized to do so, and the ownership and control of the customer.

Benefits of Effective Risk Assessment and Robust Compliance

The favorable resolution of the FCPA investigation of Morgan Stanley by the DOJ and the SEC in 2012 provides recent and powerful evidence of the benefits to financial institutions of investing resources to conduct regular risk assessments and periodically enhancing their compliance programs based on assessment results. In that matter, a former Morgan Stanley managing director pleaded guilty to one count of conspiring to circumvent the system of internal controls that the bank maintained to prevent violations of the FCPA. Both the DOJ and SEC, however, declined to charge Morgan Stanley itself, citing the company's effective ethics and compliance program as the primary basis for this decision. Morgan Stanley's internal policies, "which were updated regularly to reflect regulatory developments and [the company's] specific risks," were underscored in press releases and subsequent comments by U.S. officials as a key reason for the determination not to prosecute the bank.

UNDERSTANDING BANKING FRAUDS AND PREVENTING THEM

Below are some of the top areas in the Fintech space where fraudulent activities are on the rise.

- Forging documents and credit scores to avail big loans.
- Purchases using stolen credit cards, debit cards, passwords, etc.
- Telephone Call Scams to steal account details.
- Fraudulent Payment transactions.
- Money Laundering.
- Forging account statements for tax and loan benefits.

Whenever fraud is committed, the loss is not only born by the victim who was exploited, but the reputation of the financial institution involved in the fraudulent transaction also takes a hit. There is also a hefty penalty charged on the banks and NBFCs by financial regulators for allowing their platform to be used for fraud.

1. Automated Bank Statement Analysers during KYC

The best time to prevent fraudsters from exploiting the platform is to screen them well during onboarding itself. Most of the banks and NBFCs ask for bank statements of the potential borrowers at the time of onboarding. Analysing the bank statement in detail gives a lot of information about the customer's financial behaviour.

Traditionally banks employ hundreds of employees to assess a borrower's income and spending patterns based on the transactions available in their bank statements. But manually going through pages of bank statements is tedious and prone to human errors or, at times, even subjective biases.

This is where the automated bank statement analysers, called BSA engines, come to the rescue. A lot of Fintechs are now involved in creating intuitive and dynamic rule-based automated BSA solutions.

Below are some of the parameters an automated BSA can look for in the bank statements. All these are characters of fraudsters that need to be analysed further by the banks and NBFCs.

- Cheque Returns
- Charges and Penalties
- Circular Transactions
- Multiple smaller credits
- Frequent Cash Deposits
- Frequent Cash Withdrawals
- Equal Debits and Credits
- Disproportionate Foreign Transactions

2. Leverage Digital Footprint data during customer screening

Traditionally, the products used by banks and NBFCs for customer screening check the creditworthiness of the potential borrower by accessing a person's CIBIL score from the credit bureaus. When a borrower has a good financial record or CIBIL score, that can be used as a reliable indicator of one's creditworthiness.

But there is a massive unbanked population in India, who do not have much financial history. New borrowers can leverage digital footprint data as almost everyone carries a smartphone today and carries out digital transactions.

Fintech products used for customer screening can use enhanced technologies like data mining, analysis, and machine learning to use the digital footprint data and compare it with a fraudster's behaviour to arrive at a more accurate creditworthiness score.

3. Prevent payment frauds using multi-factor authentication and biometrics

With the wide variety of banking channels like physical, telephonic, online, mobile, ATM, POS, etc., banks can no longer have a simple user id and password as the authentication technique.

Making payments and purchases using stolen credit cards, debit cards, or online banking credentials is getting popular. Even though most payment systems use additional authentication like One Time Password (OTP) for every debit transaction, it is not always fraud-proof.

In case the mobile phone is also stolen, it becomes easy for the fraudster to commit a financial crime. This is where digital payments are playing the role of a game-changer. With digital wallets, there is no need to expose one's credit and debit cards to anybody. Digital payment systems can also use advanced biometrics that cannot be forged to authorise the transactions in addition to the regular authentication methods.

4. Advanced Transaction Monitoring and Instant Notification

Banks and NBFCs can use Fintech products built using AI and Machine Learning algorithms to trace customers' transactional behaviour. If any transaction is abnormal, the transaction is to be allowed only after additional authorisation from the customer.

This way, even if a fraudster uses a customer's stolen identity to do a transaction, based on the nature of the transaction, geography, time, and the amount involved, the transaction engine can trigger additional authorisation before allowing the transaction.

5. Balance Sheets and Tax Statement Analysers for SMEs and Corporates

When corporates and SMEs seek business expansion loans, the main documents any bank or NBFC seeks are the balance sheet and tax filing documents. Since the balance sheet formats are almost the same for all registered firms, Fintech can build smart products in lines of BSAs, scan through the balance sheet, tax filing, and look for any anomalies.

Since this process needs meticulous analysis, this can be done flawlessly and efficiently by an automated tool. This way, banks can identify any forged balance sheets quickly.

6. Avert Phishing Attacks by Building Firewalls in email and text servers

Phishing attacks (using fraudulent communications through emails, text messages, etc.) are commonly used by fraudsters to steal customers' financial information. Fintechs can build advanced firewalls that can be integrated with email inboxes and text message boxes. They scan through emails and alert the customers whenever they click on a link from the emails.

In conclusion, fraud detection and prevention is a continuous process. Like how the fraudsters keep coming up with newer ways to outsmart the financial products, the Fintech industry is also evolving with the introduction of newer products using advanced technologies like Artificial Intelligence, RPA, etc., to prevent and detect frauds.

As technologies evolve, financial products that consume these technologies are also expected to improve and become more efficient to the financial industry to build a digital fortress that is safe and secure from fraudsters.

Recovery Management

India is facing a growing problem of non-performing assets (NPA) which is dangerous as it is a symptom of an ailing banking sector which in-turn affects the economy of India as well. In the Financial Year of 2017 the value of Non- performing assets for the private sector banks stood at approximately 940 billion rupees.

This figure almost doubled in 2018 as the value of NPA's was at approximately 1.86 trillion rupees and in the financial year of 2019 this figure stood at approximately 1.84 trillion rupees[1]. In 2020 the problem is set to get even worse as a report of the RBI revealed that the gross non-performing assets ratio (which is the figure acquired by dividing net NPA by the total advances given by the bank) of commercial banks could worsen to as high as 14.7 per cent at the end of financial year 2020-21, from 8.5 per cent in March 2020, if the economic impact of the pandemic is severe.

The gross non-performing assets ratio is a measure of the overall quality of a bank's loan book and this report clearly shows that the banks could get even more stressed resulting in more problems and hardships for the banking sector. The increase in the NPAs has many adverse effects on the banking institutions as it impacts the bank's profitability, return on assets, net interest margins etc. and it also has an impact on the flow of credit as well as the growth of the economy as a whole. Therefore, it is important to understand what exactly a Non-Performing Asset is, how it works, the impact it has and the steps taken by the government to solve this problem.

What led to the rise in NPAs?

Some of the factors leading to the increased occurrence of NPAs are external, such as decreases in global commodity prices leading to slower exports. Some are more intrinsic to the Indian banking sector.

A lot of the loans currently classified as NPAs originated in the mid-2000s, at a time when the economy was booming and business outlook was very positive. Large corporations were granted loans for projects based on extrapolation of their recent growth and performance. With loans being available more easily than before, corporations grew highly leveraged, implying that most financing was through external borrowings rather than internal promoter equity. But as economic growth stagnated following the global financial crisis of 2008, the repayment capability of these corporations decreased. This contributed to what is now known as India's Twin Balance Sheet problem, where both the banking sector (that gives loans) and the corporate sector (that takes and has to repay these loans) have come under financial stress.

When the project for which the loan was taken started underperforming, borrowers lost their capability of paying back the bank. The banks at this time took to the practice of 'evergreening', where fresh loans were given to some promoters to enable them to pay off their interest. This effectively pushed the recognition of these loans as non-performing to a later date, but did not address the root causes of their unprofitability.

Further, recently there have also been frauds of high magnitude that have contributed to rising NPAs. Although the size of frauds relative to the total volume of NPAs is relatively small, these frauds have been increasing, and there have been no instances of high profile fraudsters being penalised.

What is being done to address the problem of growing NPAs?

The measures taken to resolve and prevent NPAs can broadly be classified into two kinds – first, regulatory means of resolving NPAs per various laws (like the Insolvency and Bankruptcy Code), and second, remedial measures for banks prescribed and regulated by the RBI for internal restructuring of stressed assets.

The Insolvency and Bankruptcy Code (IBC) was enacted in May 2016 to provide a time-bound 180-day recovery process for insolvent accounts (where the borrowers are unable to pay their dues). Under the IBC, the creditors of these insolvent accounts, presided over by an insolvency professional, decide whether to restructure the loan, or to sell the defaulter's assets to recover the outstanding amount. If a timely decision is not arrived at, the defaulter's assets are liquidated. Proceedings under the IBC are adjudicated by the Debt Recovery Tribunal for personal insolvencies, and the National Company Law Tribunal (NCLT) for corporate insolvencies. 701 cases have been registered and 176 cases have been resolved as of March 2018 under the IBC.

Loss Risk Forecasting

What is Risk:

When most investors think about investing, they first think about risk. While the financial markets and experts define risk as a mathematical number (e.g. volatility), common investors think of risk as possibility of losing their investment (fully or in parts) or making less than expected returns.



Globally, more banks are trying to make their operational risk management programs more forward-looking. Banks should seize the opportunities today's advanced tools and vast data pools make possible. Predictive analytics techniques, machine learning, and artificial intelligence can help efficiently build and mine large and complex data sets that combine traditional Basel operational risk loss data with other data sources, including transaction data, non-transaction data, and external data.

These aggregated data sets provide billions of data combinations that can drive vastly improved risk identification methods through analytical results and insights. The data combinations can also greatly increase the likelihood of uncovering patterns and correlations that previously weren't noticed until it was too late—if ever. This can help an organization prevent unpredictable outcomes and reduce operational losses and capital impacts.

Today, risk is dynamic, spread across the globe and has greater, more far-reaching impact.

Managing financial risk, optimising capital and enhancing your yields means more than simply balancing the books. Companies are exposed to an increasingly complex set of financial risks through geopolitical instability, exchange rate volatility, international and local regulations, relationships with customers and other stakeholders and variations in the capital markets. At the same time, investors demand that companies deploy capital efficiently, drive improvements in cash flow, manage risk to establish financial stability and maximise value for stakeholders.

Analytics can help CROs make more intelligent, data driven decisions to mitigate enterprise risk. Marketing officers are always under pressure to increase their share of wallet, derive higher ROI from their marketing campaigns and evolve with the rise in digital technology to offer an enhanced customer experience.

A loss forecasting service can be very cost-effective. Every company has some risk while doing business, and it has become essential to analyze the risk before taking any decision. Analyzing the threat has become crucial for large companies for their strategic decision making and is called Risk Analytics. For Business intelligence and data science in Finance, Risk Analytics has become vital areas.

A company can increase the security and trustworthiness of the company using risk analytics of data science. Data is the core of Risk Management and Risk Analysis because it measures the gravity of the damage and multiplies it with the frequency of loss. The knowledge of problem-solving, statistics and maths is essential in the field of Risk Management for any professional.

Raw Data majorly consists of unstructured data which cannot be inserted into a standard excel spreadsheet or a database. Data science plays a significant role in this using their frameworks to analyze the data. A company faces various kinds of risk which can originate from the market, credits, competitors, etc. The first step in managing the risk is identifying the threat. After that, monitoring and prioritizing the risk is essential.

A company can use massively available data like financial transactions and customer information using which they can create a scoring model and optimize the cost. This is an essential aspect of risk analysis and management which is used to verify the creditworthiness of a customer. Many companies now employ data scientists to analyze the creditworthiness of customers using machine learning algorithms to analyze the transactions made by customers.

Accurate, timely loss forecasting, when coupled with prompt, appropriate adjustments to operations, will produce a more financially successful, stronger, stable organization. This can lead to higher profits and bigger employee bonuses!

Now risk managers can partner with operations and HR, because Loss Forecasting provides a benchmark for successful process and procedural changes.

A Risk Manager, – and other people responsible for managing an organization’s risks should have significant knowledge of that organization’s vital information on a timely basis. Otherwise, too many things simply fall through the cracks, creating lower profits or needless losses.

Required Data:

For proper Risk analysis, a minimum of 4 years of historical loss and exposure data is required.

Risk Profiling

What is Risk profiling:

An investor makes investments in order to achieve certain financial goals. A risk profile helps an investor understand how much risk they can take vs how much risk they should take to achieve their goals.

Risk profiling involves assessment on four fronts:

1. Risk capacity: Level of financial risk that an investor can manage comfortably based on his life situation (e.g. risk capacity will be higher for a young salaried investor vs a middle aged man with two children).
2. Attitude to risk: Investor understanding of the concept of risk and how it applies to their financial life.
3. Risk tolerance: Investor’s ability to cope at psychological level with the volatility of capital markets. (e.g. response/reaction towards market corrections.)
4. Risk requirement: Investor’s financial goals to understand where they are headed and their current resources to identify risks they may be required to take up to achieve certain goals (e.g. invest in equities to plan for retirement).

How is risk profiling done:

There are multiple risk profiling tools that are available online and you can use them to get your risk assessment done. Upon completion you will be put into one of the risk buckets depending on your responses to the given questions (e.g. low risk taker, medium risk taker, high risk taker). Once you know your risk profile you can determine what kind of investor you are, what kind of returns you should expect from your investment portfolio and what kind of investment portfolio you should have.

How profiling helps a beginner investor:

A good financial plan balances your goals with your capacity to take investment risks and your risk tolerance. As a beginner investor, knowing your risk profile can benefit you in multiple ways:

- Take the right risk – as per your capacity and requirement.
- Tapping suitable investment opportunities – get to know the right investment asset mix depending on your risk profile (e.g. appropriate balance of stocks, bonds, derivatives, different types of mutual funds, etc).
- Readiness to manage surprises in returns – you are more aware of how you may react and how best to keep your emotions from getting in the way.

Thus, Risk profiling is extremely useful for all investors, but even more so for beginners, as it helps set the right investor expectation and gives an excellent opportunity for advisors to get a glimpse of their client's aspirations, attitude and tolerance.

Credit Risk Management & Predictive Analytics

Lending is becoming more future-oriented and Predictive Analytics can help financial institutions be at the forefront of innovation. All types of credit risk management require data analytics, and increased data availability and processing tools will bring new credit risk management opportunities. Predictive analytics is the practice of deriving information from existing data in order to identify the likelihood of patterns and predict future outcomes and trends. It forecasts what might happen in the future with an acceptable level of reliability and incorporates what-if scenarios and risk assessment.

Recognized by Gartner, CRIF's expertise in predictive analytics is shown by the development of various scoring projects in many including Bureau scoring models, spanning over 18 countries which in total are used to make hundreds of millions of score calculations and decisions each year around the world.

Rating systems are a core competency in CRIF, thanks to CRIF's Rating Agency experience, we provide rating model development from estimation, validation, and review to calibration and evaluation of economic groups.

Credit Risk Score Development

- **Data management** to extract value from data businesses require solutions to help them extract, align and distill what's essential and quickly determine analytical interpretations.
- **Scoring models** that permit optimisation of any financial institution process, which they are developed for. CRIF provides a full portfolio of modelling tools and expertise, empowering business analysts, from beginners to advanced modelers, to develop, build, test, deploy and manage predictive models. Types of scorecards we provide are:
- **Application Scorecard:** These are tools that allow organisations to predict the probability that an applicant will behave in a certain way, helping businesses to make effective automated decisions. Application scorecard for credit assesses the likelihood of default which means it predicts the risk of a customer paying a or not. In the credit risk application scorecard, the output is usually a numeric score provided for each applicant, with higher scores corresponding to lower levels of estimated risk. This supports lenders to make accurate and consistent decisions on whether to approve, review or decline applicants. Application scorecards can also help predict many other different metrics such as an applicant's affordability (ability to pay), potential future profitability and the likelihood to churn (attrition) etc.
- **Behavioral Scorecard:** Do you know who are your most profitable customers are? Are your customers defaulting in their payments with you or other lenders? Behavioral scorecards help in identifying, retaining and growing the right customers for our

businesses. These quantify the customer behavior to improve credit portfolio management and customer management. With CRIF's behavioral scorecards lenders can make more customer-centric decisions, respond effectively to their individual needs, enhance control of risk exposure, create a more effective pricing program and accurately target current and prospective customers for cross-selling programs.

- **Collections Scorecard:** These scorecards facilitate debt management decisions. By considering the past behavior, it identifies the risky customers. Appropriate treatments can then be initiated at the earliest on the customers based on their risk levels to protect the business assets with the most cost-effective mechanism applicable. Collection scorecard plays a significant role in the profitability of the business by minimizing the credit losses. This enables the business to minimize the provisions taken against the credit. The provisions directly impact the capital allocations that could otherwise be invested in the growth of the business. CRIF's collections scorecard can help organizations prioritize collection efforts, minimize defaults, maximize recoveries and reduce overhead costs by identifying customers who have a higher propensity to pay and targeting them with innovative and tailor-made collection strategies.

Portfolio Stress Testing

A stress test, in financial terminology, is an analysis or simulation designed to determine the ability of a given financial instrument or financial institution to deal with an economic crisis. Instead of doing financial projection on a "best estimate" basis, a company or its regulators may do stress testing where they look at how robust a financial instrument is in certain crashes, a form of scenario analysis. They may test the instrument under, for example, the following stresses:

- What happens if unemployment rate rises to v% in a specific year?
- What happens if equity markets crash by more than w% this year?
- What happens if GDP falls by x% in a given year?
- What happens if interest rates go up by at least y%?
- What if half the instruments in the portfolio terminate their contracts in the fifth year?
- What happens if oil prices rise by z%?

This type of analysis has become increasingly widespread, and has been taken up by various governmental bodies.

Stress testing involves running computer simulations to identify hidden vulnerabilities in institutions and investment portfolios to evaluate how well they might weather adverse events and market conditions.

Types of Stress Testing

Stress testing involves running simulations to identify hidden vulnerabilities. The literature about business strategy and corporate governance identifies several approaches to these exercises. Among the most popular are stylized scenarios, hypotheticals, and historical scenarios.

In a historical scenario, the business—or asset class, portfolio, or individual investment—is run through a simulation based on a previous crisis. Examples of historical crises include the stock market crash of October 1987, the Asian crisis of 1997, and the tech bubble that burst in 1999–2000.

A hypothetical stress test is generally more specific, often focusing on how a particular company might weather a particular crisis. For example, a firm in California might stress-test against a hypothetical earthquake or an oil company might do so against the outbreak of war in the Middle East.

Stylized scenarios are a little more scientific in the sense that only one or a few test variables are adjusted at once. For example, the stress test might involve the Dow Jones index losing 10% of its value in a week.

As for the methodology for stress tests, Monte Carlo simulation is one of the most widely known. This type of stress testing can be used for modeling probabilities of various outcomes given specific variables. Factors considered in the Monte Carlo simulation, for example, often include various economic variables.

Companies can also turn to professionally managed risk management and software providers for various types of stress tests. Moody's Analytics is one example of an outsourced stress-testing program that can be used to evaluate risk in asset portfolios.

Portfolio Stress-Testing is a holistic and fundamental approach to stress-testing. The focus is on the underlying drivers of portfolio performance and link these drivers to alternative scenarios based on a range of possible outcomes given the current business cycle. By linking performance drivers to local and global economic activity, our approach generates more realistic results to assess risk.

MARKET SHARE ESTIMATION

The starting point for estimating market size is to understand the problem you solve for customers and the potential value your product generates for them. This is an aspect that many startup founders in the innovation community tend to overlook, since they get excited about the product they've developed without thinking about how it benefits their audience. For most businesses, the concept of market sizing is readily understood but not easily accomplished. Many get stuck on establishing boundaries or defining the market before they even get to the data analysis and implications of their research. Determining market size can answer strategic questions about levels of investments in the business and profitable growth targets. Market sizing can also serve as a quick understanding of the potential for a B2B market opportunity in terms of volume or value, and is therefore pertinent to business strategy and decision making.

Depending on your technology, you may have to choose which customer problem to solve first. If this is the case, completing the exercise below may help you better grasp the market size for each application. This will make it easier to prioritize which problem to solve first.

How to Calculate a Company's Market Share

To calculate a company's market share, first determine a period you want to examine. It can be a fiscal quarter, year or multiple years. Next, calculate the company's total sales over that period.

Then, find out the total sales of the company's industry. Finally, divide the company's total revenues by its industry's total sales.

Investors can obtain market share data from various independent sources, such as trade groups and regulatory bodies, and often from the company itself.

For example, suppose you want to calculate a toy manufacturer's market share over one fiscal year. The toy manufacturer had total revenues of \$20 million, and the toy manufacturing industry had total revenues of \$200 million over one fiscal year. To find the toy manufacturer's market share, divide \$20 million by \$200 million. The manufacturer's market share is 10%.

Comparing Market Share in an Industry

Market share can also be used to compare similar companies within the same overall industry. For example, suppose another toy manufacturing business has total revenues of \$40 million. This toy manufacturer has a 20% market share of the industry. This signals that this toy manufacturer out-competes the toy manufacturer from the previous example.

It is also possible to use market share over multiple periods to see how well a company fares against its competitors and whether the company is growing.

Companies are always looking to expand their share of the market, as well as grow the size of the total market by appealing to larger demographics, lowering prices, or using advertising.

UNIT V : Applications of Business Analytics for Customer and Market Analytics

Customer Analytics and Intelligence helps organizations understand their customers, reduce attrition, increase loyalty, enhance customer life time value and maximizes profitability. Market Equations India offers impact-based Customer Data Analytics and Intelligence outsourcing services to organizations globally helping them enhance customer engagement and experience leading to increased customer retention, loyalty and adoption. Our result-oriented Customer Data Analytics and Intelligence outsourcing services help organizations have a clear and precise understanding of their customer base and their changing demands and expectations leading to improved customer acquisition, retention, loyalty and life time value.

Customer analytics insights backed by high quality predictive models takes customer engagement to a whole new level by influencing their next set of actions that has a positive reflection on business performance.

First thing in business is to understand your customers. The customer is always right. Decision makers understand this but fail to implement a integrated customer experience strategy leading to high customer attrition and diminishing profitability. Market Equations helps organizations leverage analytics to answer these key questions:

- What happened,
- Why it happened,
- What will happen next and
- What action to take to get the best outcome.



LOYALTY ANALYTICS

What is Customer Loyalty?

It shows customer's desire and inclination to buy products from a brand again and again. This happens when a customer has had a positive experience, has been satisfied with the value of the product as well as from the services he/she gets the purchased products.

In order for customers to continue purchasing your brand, entrepreneurs always work on providing good customer services.

Loyal customers tend to spend more on brands that they like and also have the tendency to share their positive experience with their family and friends. 'Word by mouth' a new marketing technique has gained importance as it proves to be more effective to grab the attention of new customers rather than advertising your brand.

Grabbing the attention of "new customers" and trying to convert them into "loyal customers" is a tedious, time consuming and an expensive task. Instead simply by having a word by mouth about your brand can attract new customers. But how can this be done? How do you use positive reviews, eye-catching tweets Instagram mentions and other social media platforms to boost the development of your business?

It's over here where the customer loyalty program proves to be useful.

What is Customer Loyalty Program?

It is basically an organized collection of customer data which is often used to offer customers add on privileges and benefits or services to loyal customers in the form of reward points, coupons vouchers or any other such benefits.

Customer Loyalty Programs are needed for:

- Rewarding the customer for being special.
- To create a bounce in enrolments/memberships
- Grabbing the eyeballs of new customers to new attractive offers
- To generate and show a keen interest in existing customers

In India customer loyalty programs have been active since 1995. Every major vertical today be it travel, hospitality, healthcare, retail, telecom, media outlets etc. have invested in some kind of loyalty programs. The biggest challenge faced by Indian marketers is deriving actionable insights from customer data. Marketers lack in knowledge such as how to develop, administer and use loyalty programs in an effective manner.

In India customer loyalty programs are still in its embryonic stage as most of the programs are either point driven or offer discount and voucher schemes. It is tough to benefit from a CRM Program without well-organized database.

Why does customer loyalty matter?

Customer loyalty is a strong predictor of a businesses' long-term viability because loyal customers are durable. Unlike customers that feel neutral or even negatively about a product or service, loyalists actively want to remain customers. They're more resistant to competitors' marketing and more likely to purchase again, purchase additional products, and tolerate errors such as outages or irritating support interactions. Businesses with a loyal customer base often enjoy:

- More repeat purchases
- Higher retention
- Higher average order value (AOV)
- Higher engagement
- Forgiveness for poor service
- More vocal customer advocates

Of all the benefits of customer loyalty, advocacy is among the most valuable. Loyal customers often share their experiences online and with friends, and their recommendations typically carry more weight than recommendations from brands. "Today's buyers simply don't trust your company's slick marketing messages," says Chris Newton, VP of Marketing for the advocacy software Influitive. "Instead, they seek out recommendations before purchasing. Buyers trust the authentic words of their peers. That's why you need to find a way to systematically improve your customer loyalty so they spread their love of your brand. Marketing messages are more meaningful coming from advocates, and they're much more likely to influence your prospects." By increasing loyalty, businesses activate their customers to market their products more credibly for little or no additional cost. To consistently earn customers' loyalty and get advocates talking, however, it helps for teams to first define 'loyalty' and learn to calculate it.

Customer Loyalty Analytics

Each and every customer contributes to the growth and success of your business. But from a loyalty program perspective, they're not all of equal importance.

Some customers in your program are far more valuable than others — in fact, it's typical for around 20% of members to drive up to 80% of future profits. Customer loyalty analytics help you identify this key minority, allowing you to take actions today that maximize the likelihood that you'll unlock their value.

Analytics also enable you to go beyond the current high-value members and identify those with the potential for future increases in customer value. Spotting these members early on enables you to plan your marketing budget more effectively to drive up program ROI.

Once these key members have been identified, the way you target members also becomes informed by your loyalty analytics. What elements of the rewards program are they most attracted to? What discounts, services or options are they most likely to utilize? A well-built customer loyalty analytics tool will guide targeting efforts in the right direction.

How Analytics Is Transforming Customer Loyalty Programs

How do you determine if your loyalty program is working well? Use data to steer your customer loyalty program in the right direction.

McKinsey found that “executive teams that make extensive use of customer data analytics across all business decisions see a 126% profit improvement over companies that don’t.” By instituting a loyalty program, you not only improve customer appreciation of your business, but you also increase the chances that existing clients will share this joy with those close to them,

- Focusing on Retention**

One primary mission of loyalty programs is to increase customer retention. You want buyers to remain with your brand after they make a purchase. For your business, higher retention means a steady flow of revenue. And it cuts down on your costs to constantly acquire new customers.

Therefore, your loyalty programs must be effective. They need to serve a real purpose for the consumer, not just your bottom line. To provide the best customer experience, fuse data into your retention strategies. It will impact how your team approaches the buyer.

“Influencing customer loyalty in this way doesn’t require magic, it requires data – usually data that you already have but aren’t using to full advantage. Regardless of industry, most organizations today generate mountains of data.”

Uncover the correlation between customer characteristics and purchasing behavior. Assign your team to analyze the current data of your most valuable customers. And learn which characteristics these customers have in common and which traits are dissimilar.

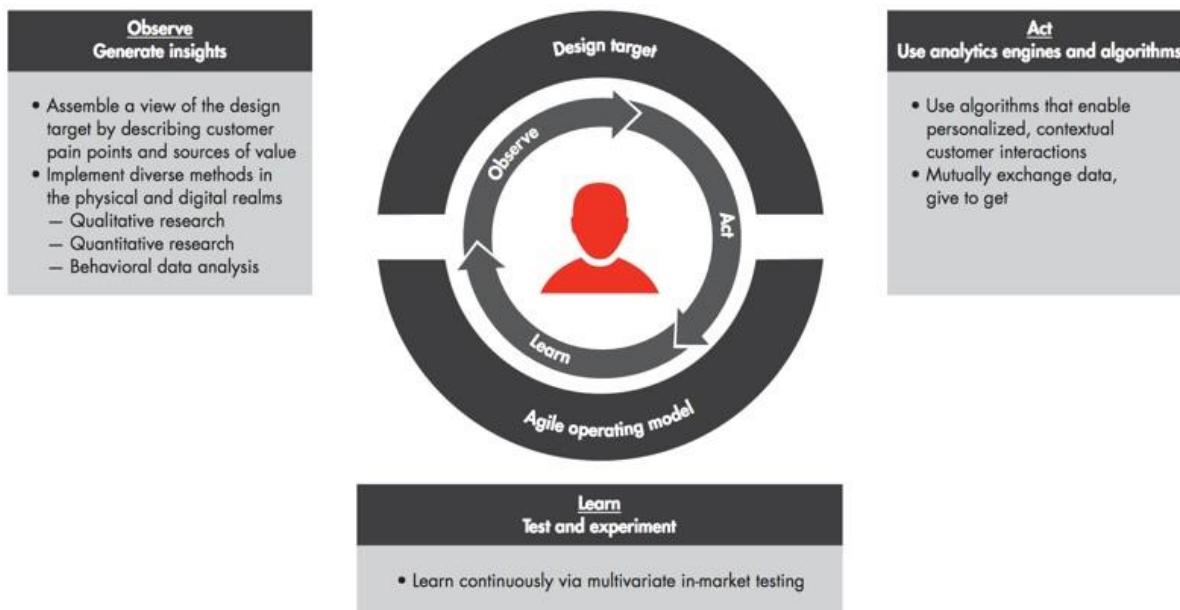
Consider data an ongoing process of observing, acting, and learning. Improve your loyalty programs by taking action on your insights. Measure success by monitoring your customer lifetime value, loyal customer rate, and redemption rate.

Start with retention. And let the data guide you to customer loyalty.

- Targeted Product Recommendations**

Research shows that “customers that are actively engaged with brands and their loyalty programs make 90% more frequent purchases, spend 60% more in each transaction and are five times more likely to choose the brand in the future.”

Figure 2: Analytics teams should generate insights that will help improve the customer's experience



Source: Bain & Company

Sending targeted product recommendations is one way to keep customers engaged. Because if they are not receptive to certain products, consumers will feel more inclined to take their business elsewhere.

Integrate real-time purchase data with historical purchase data to make specific recommendations. For example, if a small business bought payroll software from you, their team might be interested in purchasing your series of on-demand accounting webinars.

“Consumer data must be analyzed to create highly targeted product recommendation offers. Analyze consumer data such as demographics, lifestyle, products purchased by category and type, frequency of purchase, and purchase value,” states Larisa Bedgood, director of marketing at DataMentors.

It’s key not to draw wild conclusions from one piece of data. Just because a Mumbai resident buys a winter coat doesn’t mean he wants to be flooded with similar recommendations. The consumer might have bought it as a gift for a friend living in Manali.

So, gather multiple data points in order to make intelligent recommendations. You don’t want to frustrate loyal customers.

Your brand also can take a different approach. Use social proof to your advantage. If consumers are hesitant about particular products, remind them that other people are buying the product, too.

Home Depot uses this tactic by displaying a list of bestselling inventory. It persuades the customer to join the crowd.

- **Timely Promotions**

For customers, loyalty takes effort. They receive lots of promotional ads everyday to try products from other brands. Appreciating your consumer's urge to resist the hype is important. Mobile phone carriers lead the way in baiting consumers to switch their services. Airtel offers cell phone users up to Rs.650 in credit just to say bye to Vodafone, BSNL, or Jio.

To keep their loyalty, customers will hold your team accountable. They expect timely promotions that not only fit their buying habits but also their lifestyles.

At the end of the day, you want to deliver the right offer at the right time. This will increase the likelihood of the promotion redemption.

Monitor the sales data to learn when promotional codes are redeemed. Do your consumers use promotions more often in the morning? Right after a sales announcement? Or during summer months?

"By creating a time-sensitive sales promotion and having a good grasp on your target customer demographic, you'll be able to incentivize the right actions, get them to respond, and grow your business in the process,"

Moreover, analyze your reports to discover the best product promotions. A timely discount matched with the wrong product won't be useful for the consumer or your company.

Segment your customers to offer relevant discounts for multiple channels—in-store, online, and mobile. Every loyalty member doesn't have to receive the same offer.

For instance, Starbucks offers its Gold members the opportunity to earn double stars. The coffee company surprises its loyal consumers on a different day each month. This technique increases the excitement and prepares customers to spend more money on a particular day. Don't wait for your competitor to offer your customers a good deal. Start creating your own timely promotions.

- **Personalized Rewards**

Everyone likes to be rewarded. It signifies that you've done something commendable. And incentives compel you to continue the rewarded behavior.

Recognize the value of your customer's actions. Because that's what you're rewarding.

You can offer perks based on monetary transactions, shopping frequency, or even survey responses. It's all about showing appreciation for consumers' actions.

But it's your team's job to appropriately reward customers. Don't expect people to buy \$1000 worth of services in one month if your highest service retails at \$10.

In addition, manage your loyalty members' expectations. They shouldn't expect your brand to give away free Beyonce tickets every day.

Personalized rewards ensure you're giving your customers what they desire. It also shows that you are truly invested in the customer experience.

Send a simple email survey asking consumers what types of incentives excite them. Or conduct social media listening to identify useful prizes that can make your customers' lives better.

Decathlon Sporting Goods sends emails asking customers for their opinions. The company uses the information to improve its inventory and customer service.

Remember to focus on maintaining positive relationships with your consumers. Because that's the ultimate goal for loyalty initiatives.

You want people to feel comfortable with your brand. Aim to offer rewards that bridge the gap between the consumer-brand relationship.

"A significant aspect of customer loyalty comes down to your likability. People will almost always remain committed to a brand if they believe they've developed a genuine and mutually beneficial relationship.". Tailor your rewards to satisfy your customers. Offer them something special.

- **Analyze Customer Loyalty**

Customer loyalty can lead to retention. That's why your team must use data to drive your loyalty programs.

Give consumers targeted product recommendations they can't resist. Send promotions at the right time. And personalize rewards so the customer feels part of the brand.

Look at the data. Improve customer loyalty programs.

CUSTOMER LIFE TIME VALUE

The lifetime value of a customer, or customer lifetime value (CLV), represents the total amount of money a customer is expected to spend in your business, or on your products, during their lifetime.

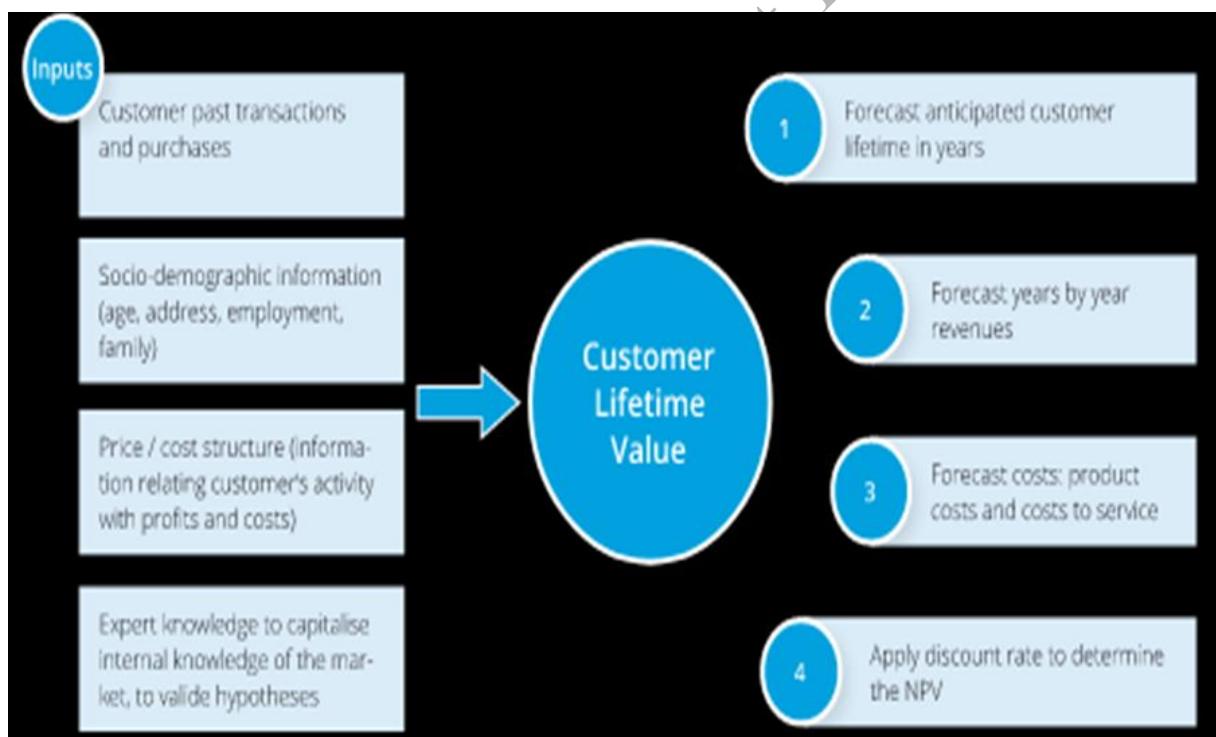
The world of business has completely changed and revolves around its customers more than ever. The customer-centric approach is the new norm in today's market. The reason for that is the ample choices people have when choosing a product/service.

In this era of businesses fighting against each other to better serve and seize customers from their competitors, the need for them to grow and retain their existing customer base is very important.

But similar to the process of acquiring customers, there is a huge cost associated with the process of retaining existing customers too. (by giving discounts, targeted offers, etc.)

So, you might think, do they need to retain every single customer? Well, not really. In every business, some customers create more value for the business by being a loyal customer and some are just one-time buyers. Identifying such groups of customers and targeting only the high-value customers will help the business to at least sustain in this competitive market. Now, the real challenge begins—How to find the customer value? Before answering this question, let's just define what does “customer value” means.

Customer lifetime value (CLV) is the expected value of the future relationship with a given customer. The potential value of a group of customers or of the entire customer base is called customer equity. The calculation of customer lifetime value has four components: lifetime duration, cash in, cash out and the discount rate as illustrated below.



Based on the input data, we build customer micro-segmentation and customer profitability model using homogeneous Markov chains. The CLV model can be used together with a churn model in a matrix to identify the target group eligible for a marketing action, which is usually a contact with a special offer.

Predictive CLV

So what's behind the concept of the predictive CLV? First, its aim is to model the transactional behaviors of your customers to predict what actions they will take in the future. This method is a great indicator of customer lifetime value, better than historic CLV.

By applying Customer Lifetime Value marketing managers can easily arrive at the rupee value associated with the long-term relationship with any customer. It is difficult to predict how long each relationship will last, but marketing managers can make a good estimate and state CLTV as a periodic value.

It is useful metric used by marketing managers especially at a time of acquiring a customer. Ideally, lifetime value should be greater than the cost of acquiring a customer. Some also call it a break-even point.

The basic formula for calculating CLTV is the following (1):

(Average Order Value) x (Number of Repeat Sales) x (Average Retention Time)

For example, let's say you run a Health Club where customers pay Rs 1000 per month and the average time that a person remains a customer in your club is 3 years. Then the lifetime value of each customer is (according to the formula above):

Rs 1,000 per month x 12 months x 3 years = Rs 36,000. This means each customer is worth a lifetime value of Rs 36,000.

Once we calculate CLTV we know how much the company can spend on paid advertising such as Facebook ads, YouTube ads, Google Adwords etc. in order to acquire a new customer.

PROPENSITY ANALYTICS

Today customers have so many choices and so much information at their finger tips that it becomes even more difficult to acquire or retain them. A carefully designed customer acquisition strategy can help you understand what your customers want, at what value and when helping you acquire, retain and increase wallet share. Propensity analysis offers some common approaches to find answers to the questions like ***when to promote a product and when not to in order to keep the offers personalized?***

Propensity models likelihood – response models predictive analytics

To predict which prospects are ready to make their first purchase, a likelihood to buy model evaluates non-transaction customer data, such as how many times a customer clicked on an email or how the customer interacts with your website. These models can also take into account certain demographic data. For example, in consumer marketing, they may compare gender, age, and zip

code to other likely buyers. In business marketing, relevant demographics may include industry, job title, and geography. SwiftERM automatically appreciated the uniqueness of the visiting customer capturing each and every click made as they navigate throughout the website. Nuance is all.

Here's how it works: the models compare the pre-purchase behaviour of prospective buyers to the pre-purchase behaviour of thousands or millions of previous customers who ended up buying, comparing attributes like which emails they opened and what products they spent the most time looking at. The prospects that behave most like the previous buyers are tagged as "high-likelihood buyers" and SwiftERM then alters the way it interacts with them to increase the likelihood of closing a sale. Once you're armed with this data, you have a heightened likelihood of a successful return for each prospective customer.

Predicting the Likelihood to Buy for First-Time Buyers

For consumer marketers, likelihood to buy predictions allows you to decide how much of a discount you might allocate to a certain customer because people who are already more likely to buy won't need as aggressive of a discount as customers who are less likely to buy. The models then get better over time, as companies collect more data and automatically test whether predictions actually become reality.

For instance, the large Indian household appliance manufacturer Whirlpool maintains a call centre where employees are given a list of customers who are likely to be ready to buy a new washing machine within the next few months. Agents then make calls to these customers with offers such as a year of free detergent with the purchase of a washing machine. The tactic works well for considered purchases, such as refrigerators or cars, and larger-ticket items such as high-end fashion apparel.

A high-end shoe brand provides store associates with lists of customers to call too. The store associates have already developed strong relationships with their customers, but they can be even more successful when armed with predictive analytics. Employees can now see which customers are likely to be interested in a certain style when a new season's shoe comes out, based on customers' past behaviour or how similar their purchase habits are to other customers. Employees can then reach out to customers with that information. A call could go something like this: "Hi Joe, it has been a while since we've spoken. I just wanted to let you know that there is a new cross-country running shoe I think you might like. It's similar to the shoes you bought two years ago but in a new material. I have put a pair aside for you in your size. If you have time, perhaps you could stop by on your way home from work to have a look?" Who would not want to receive a call or an email like that from their personal shopper?

As reported by the New York Times and others, President Barack Obama used propensity models, specifically propensity to vote for the Democratic Party, to help him win reelection in 2012. His staff of volunteers could not possibly meet with every voter in the country so the challenge was to find the undecided voters. There was no point spending time or money trying to woo diehard Republicans who would not change their minds anyway, or diehard Democrats who were already likely to vote for Obama. Rather, using propensity models, Obama's team of data scientists found

those voters who were undecided but could still be persuaded. They then focused on finding already strong Obama supporters in the undecided voter's social circle and asked them to spend time with the undecided voter to explain their views.

Predicting Likelihood to Buy for Repeat Buyers

What good is spending money to acquire new customers if they only buy once and do not return? Based on a customer's propensity to purchase, it is not only important to predict likelihood to buy for first-time buyers, but it is equally important to predict likelihood to buy for repeat buyers. Your goal is to keep customers coming back time and time again. It is happy and loyal customers who have a large lifetime value, and many customers with a large lifetime value make for large revenues and profits for your company.

Predicting the likelihood to buy for repeat buyers is a lot easier than predicting the likelihood to buy for first-time buyers because there is a lot more information to go on. The likelihood to buy model for repeat purchases evaluates earlier transactions as well as other interactions similar to the model for prospects. However, the added information derived from the first purchase can significantly improve the accuracy of the likelihood to buy model for repeat purchases, as compared to a similar model for prospects. Unlike the first purchase predictions, repeat purchase predictions utilize all interactions of the customer, such as past purchases and returned purchases. That SwiftERM adds data captured on repeat visits satisfies an enormous provision for more accurate returns for each consumer.

Because the replacement and delivery cycles for vendors, deals, services, and products can take a long time, most marketers are hyperfocused on acquiring new customers, rather than getting existing customers to come back, where the likelihood to buy first purchase models take greater importance, and the cost to keep them is infinitely lower.

Predictive models are not the only way to prioritize prospects for business marketers. However, predictive models are by far the most accurate and relatively easy to use.

For consumer marketers, likelihood to buy predictions allows you to decide how much of a discount you might allocate to a certain customer because people who are already more likely to buy won't need as aggressive of a discount as customers who are less likely to buy. The models then get better over time, as companies collect more data and automatically test whether predictions actually become reality.

For instance, the large European household appliance manufacturer Arcelik maintains a call centre where employees are given a list of customers who are likely to be ready to buy a new washing machine within the next few months. Agents then make calls to these customers with offers such as a year of free detergent with the purchase of a washing machine. The tactic works well for considered purchases, such as refrigerators or cars, and larger-ticket items such as high-end fashion apparel.

A high-end shoe brand provides store associates with lists of customers to call too. The store associates have already developed strong relationships with their customers, but they can be even more successful when armed with predictive analytics. Employees can now see which customers are likely to be interested in a certain style when a new season's shoe comes out, based on customers' past behaviour or how similar their purchase habits are to other customers. Employees can then reach out to customers with that information. A call could go something like this: "Hi Joe, it has been a while since we've spoken. I just wanted to let you know that there is a new cross-country running shoe I think you might like. It's similar to the shoes you bought two years ago but in a new material. I have put a pair aside for you in your size. If you have time, perhaps you could stop by on your way home from work to have a look?" Who would not want to receive a call or an email like that from their personal shopper?

As reported by the New York Times and others, President Barack Obama used propensity models, specifically propensity to vote for the Democratic Party, to help him win reelection in 2012. His staff of volunteers could not possibly meet with every voter in the country so the challenge was to find the undecided voters. There was no point spending time or money trying to woo diehard Republicans who would not change their minds anyway, or diehard Democrats who were already likely to vote for Obama. Rather, using propensity models, Obama's team of data scientists found those voters who were undecided but could still be persuaded. They then focused on finding already strong Obama supporters in the undecided voter's social circle and asked them to spend time with the undecided voter to explain their views.

Predicting Likelihood to Buy for Repeat Buyers

What good is spending money to acquire new customers if they only buy once and do not return? Based on a customer's propensity to purchase, it is not only important to predict likelihood to buy for first-time buyers, but it is equally important to predict likelihood to buy for repeat buyers. Your goal is to keep customers coming back time and time again. It is happy and loyal customers who have a large lifetime value, and many customers with a large lifetime value make for large revenues and profits for your company.

Predicting the likelihood to buy for repeat buyers is a lot easier than predicting the likelihood to buy for first-time buyers because there is a lot more information to go on. The likelihood to buy model for repeat purchases evaluates earlier transactions as well as other interactions similar to the model for prospects. However, the added information derived from the first purchase can significantly improve the accuracy of the likelihood to buy model for repeat purchases, as compared to a similar model for prospects. Unlike the first purchase predictions, repeat purchase predictions utilize all interactions of the customer, such as past purchases and returned purchases. That SwiftERM adds data captured on repeat visits satisfies an enormous provision for more accurate returns for each consumer.

Because the replacement and delivery cycles for vendors, deals, services, and products can take a long time, most marketers are hyperfocused on acquiring new customers, rather than getting existing customers to come back, where the likelihood to buy first purchase models take greater importance, and the cost to keep them is infinitely lower.

Predictive models are not the only way to prioritize prospects for business marketers. However, predictive models are by far the most accurate and relatively easy to use.

Essentially, propensity modeling correlates customer characteristics with anticipated behaviors or propensities. It tracks buying habits as well as other actions such as a customer's propensity to open a marketing email, sign up to a loyalty program, or participate in feedback surveys.

Its success is underpinned by the quality of your customer data and how effectively it's segmented. Say you're a national retailer with physical stores and online channels. You have three customer segments defined by their shopping frequency and spending. Namely, the frequent shoppers, the slow-and-steady customers, and the at-risk customers.

Applying a propensity modeling predictive tool to each of these customer segments will allow you to develop a far more successful, long-term sales strategy—one that responds to growth opportunities with proactive and timely cross-selling and upselling campaigns.

For example, what is the retention probability of your frequent shoppers? Is the frequency between their shopping trips, or the amount of money they spend on each shop, increasing or declining—and if so, why? Why do your frequent shoppers prefer to shop with you, and how can you leverage this knowledge to influence your slow-and-steady and at-risk customers?

Armed with this knowledge, you can pursue far more effective selling techniques that build better customer relationships. You'll be able to track how certain marketing and sales communications influence buying behavior over time and deliver highly targeted and personalized messages that hit home. Ultimately, your customer retention and acquisition strategies will see a faster and greater return on investment—saving your brand time and money in the long run.

Segmenting Your Customers

Personalized customer communications can be tricky to get right for brands of any size. It's all too easy to create too many customer segmentation models, breaking up your target audience into an impossibly diverse list of personas. Or, on the flip side, to create too few and use only a handful of overgeneralized stereotypes. To make the most of your marketing budget, you need to focus on the customers who are going to spend the most on your brand. If a persona is unlikely to generate profit, then you need to either drop that persona or reduce your investment in that specific group.

Another common mistake is when marketing teams use their own experiences to reinforce their brand's personas. Instead, they need to focus on analyzing and using the real customer data they have on hand.

Running Regular Data Checks

Data does not last forever. Out-of-date customer information will affect the quality of your models and end up costing you more money and time. A lot can happen to a customer in a short space of time; they could get married, move, have children, start studying, become a vegetarian—the list goes on. You have to continuously learn about your customers and update your data in real time. Every time you engage with them, use it as an opportunity to gather more information. Don't just expect personal details to be handed over willingly—you need to offer something in return to make it worth your customers' while, a value exchange that bonds them ever tighter to your brand.

Remember that not all data is valuable, so test and delete unprofitable data sets. To avoid getting into a situation where reams of useless customer information crowd out important, relevant data, it's best to determine the accuracy of all your data at the source; effective value exchanges will help.

Technology is increasingly proving its worth as a valuable tool for marketers and salespeople alike. That said, it won't do all the work for you. To get the most from the tools on offer, you need to apply some strategic thought—and really get to know your customers.

CHURN ANALYTICS

Customer churn and loyalty analysis helps you identify customers who are not satisfied, who are most likely to churn and the key churn triggers that lead them to take this action. Early identification of these churn drivers can help organizations take proactive action and build retention strategies to enhance customer loyalty.

Identifying these key (churn) drivers helps organizations design churn strategies to enhance customer loyalty by developing a customer focused churn management, retention and revival strategy. The Market Equations Centre of Excellence in India combines advanced analytics expertise and decades of consulting experience to help organizations design innovative and customer centric churn management strategies to enhance customer satisfaction, customer loyalty, retention and life time value.

Business should identify focus areas for customer churn and then quantify the associated business impact – both for short and long term. It starts with something as simple as agreeing on criteria to define what constitutes a churn, types of churn, priority focus area, inclusions and exclusions etc. For example, the next consideration is whether to focus on hard and soft churn or both. Hard churn refers to a defined event that signifies churn, for example, the closure of an account. However, this approach may be too simplistic, thereby requiring soft churn to be considered. A customer may be defined as having soft churned if he or she has not transacted with the business for a certain period. The length of time varies depending upon the nature of the industry and often on the customer's initial behavior (namely, transactional frequency). For example, a shorter period would be used if focusing on supermarket shopping (as we shop and eat every week) compared to booking a holiday.

Consideration should be given as to whether to focus on a specific customer group. The measurement may be limited to customers with the most valuable product, for example, home loans for banks, premium card holders, members of premium loyalty cards bucket or to another area such as the busiest time of year, Diwali, Christmas for seasonal businesses. No transaction by the premium or loyalty card holders during Diwali or other festive seasons might indicate first signs of losing that customer. Once the degree of churn is known, it needs to be expressed as a financial consequence to the business. So if the churn rate is x%, business will lose INR Y million.

To assess the financial impact depends on the value or profitability by product, service, and channel and/or customer segment. A number of assumptions based on customer knowledge can be made. For example, if the average tenure of a customer is 5 years, the expected future revenue can be built into the assessment.



Target Audience 1: Most valuable customers and those who are likely to churn first

Target Audience 2: After focusing on target audience 1, we can focus on low-value customers and those who are likely to churn

Key findings and analyst recommendations based on the analysis provide the deployment strategies for implementing customer retention and customer loyalty marketing initiatives to minimize and control the future impact of customer churn. Examples of the type of insight that one can expect include the best time to contact a customer based on a moment of truth and a defined demographic profile. The combination of likelihood to churn in conjunction with variables provides a powerful basis for marketing action.

Telecom operators including Vodafone, Reliance Communications, Aircel and MTS India are investing in technologies such as big data and analytics amid growth in data consumption as they look to retain subscribers and increase revenue per user.

Analytics allow telecom companies to improve their data product offerings by better understanding customer demands and usage patterns, according to the operators. The equipment and networks that provide data services generate significant a lot of data that the operators use to identify and proactively address the network issues they face. This also helps the operators plan their network expansion strategies better.

Vodafone India's company has deployed big data technology and is piloting various scenarios to get insights into the company's customers, usage pattern and needs. Both structured and unstructured data feeds have been integrated as part of the project. Insights from this are helping us in a big way.

The tools help in analysing the new generation of market segment – Z Gen or newer generation – that demands more digitised software-based services as against voice and SMS service. By using these technologies and without getting any intrusion into their privacy, we are able to know what type of services they use, what is the nature of the behaviour of these new generations, and then offer services as per their needs.

Customer retention and revenue growth are the largest issues which every telecom operator faces, and the ability to understand and foresee likely churn customers and their retention is critical for revenue management and organisational growth. Smart Analytics or Big Data based Analytics enabled them in understanding customer profile for providing value added services, which has helped in increasing revenues,

3 benefits to improve customer retention with analytics

1. Reduces cost to acquire customers

It's much cheaper to keep an existing customer than it is to earn a new one. In fact, it can be five times more expensive to attract a new customer, than to keep an existing one.

Stay best friends with your loyal customers, as they are extremely valuable. Once you know why your happy customers stay and why some leave, you can take the right measures to keep the right customers.

2. Easier upsell/cross-sell opportunities

It goes without saying, but your existing customers are much easier to market and sell to. Consider that usually, there are no huge customer acquisition costs associated with selling a new product or service to your existing customer base.

3. Facilitates sustainable growth

Keeping existing customers allows for more sustainable growth. Says Bain & Company, increasing customer retention rates by 5%, can increase profits by anywhere from 25% to 95%. It's clear that retaining existing customers makes the most business sense, but doing so isn't quite

that simple. One way many companies are finding a competitive advantage is through customer retention analytics.

5 types of customer retention analytics

Here is a quick rundown of 5 common types of retention analytics.

1. Prescriptive Analytics

Facilitates focusing on answering a specific question, and can help to determine the best future solution among a variety of options, and suggest options for how to take advantage of a future opportunity or illustrate the implications of each decision to improve decision-making. For customer retention, examples of prescriptive analytics include the next best action and next best offer analysis.

2. Predictive Analytics

This is the most commonly used method. Predictive analytics uses models to forecast what might happen in a future, specific situation. This could be next best offers, churn risk and renewal risk analysis.

3. Descriptive Analytics

Not always the best value results, and fairly time-consuming, it can still be useful for uncovering patterns within a certain segment of customers. This technique provides insight into what has happened historically and will provide you with patterns and trends to be able to investigate the detail. Examples of descriptive analytics include summary statistics, clustering and association rules used in market basket analysis.

4. Diagnostic Analytics

This technique is often used when trying to identify why something happened, such as looking into churn indicators and usage trends amongst customers. Examples of diagnostic analytics include churn reason analysis and customer health score analysis. It mainly looks to past events, focusing on causal relationships and sequences.

5. Outcome Analytics

Also known as consumption analytics, outcome analytics gives insight into customer behavior that drives specific outcomes. This approach is focused on consumption patterns and associated business outcomes. Use it to understand your customers better and learn how they are using your products and services.

5 strategies to reduce churn with data

If you're looking at using data to reduce churn and in effect improve your customer retention, we've got 5 tips up our sleeve that should help you off on the right foot.

1. Develop a data roadmap and stick to it

As many as 30% of the executives in the aforementioned Bain & Co study said that they lack a clear strategy for embedding data and analytics in their companies. McKinsey's findings show that taking an integrative approach, meaning seeing analytics as a strategic driver of growth instead of using it in silo or only as a part of IT, ultimately leads to achieving the desired result (McKinsey, 2014).

Successful companies do two things differently: First, they make use of the data they have. Second, they implement the organizational changes once they understand what the data tells them. So, you have the data – make sure you actually use it and enforce any changes needed in the business to make it happen quickly.

A good approach is to develop a data roadmap and stick to it. Steps that you take within the organisation can be to:

1. Ensure corporate KPIs are automated, scalable and repeatable.
2. Gather key stakeholders and define the top 3 business problems you want to solve.
3. Categorize the issues into data vs. systems issues (often you'll find that the issue is not with "data" at all, but with how people use it or manage it).
4. Prioritization of tasks is required along with assessing the technical feasibility of your plan.
5. To stay on track, reassess progress every 3 months.
6. The human factor – ensure behavioral change

Another key factor is hiring senior executives who take a hands-on approach to customer analytics. Not only do they need to understand the importance of analytics but also have the skills to analyze it themselves, so use this as a benchmark when hiring.

Although 70% of companies have data strategies in place, many will fail to deliver what's needed due to one factor alone: people. You may have the most advanced tools and excellent data scientists; however, all efforts fail without the correct behavioral changes needed internally to ultimately take action (Bain & Co 2017).

Employees may not be committed to using data analytics, internal teams may not be communicating with each other, or the data solutions adopted aren't user-friendly. Behavioral change, continuous monitoring of results, along with a "one-team approach" is needed to ensure that advanced analytics within an organisation can survive and prosper (Bain & Co, 2017). No surprises here, behaviour change being the hardest part of any performance improvement plan and why as many as 38% change efforts fail (Bain & Co, 2016).

2. Only focus on high-quality leads

Customers are less likely to churn if they are similar to your primary target customers. If you have access to data about both your customers and a list of potential customers, this is a great opportunity to focus on only those who are less likely to churn.

3. Use machine learning methods to create predictive models

Companies analyze data using different types of analytics, including predictive analytics, which is used to look at the relationships among different metrics.

To create solid customer retention strategies, we can use predictive analytics to make predictions about the future, by looking at historical data, to learn what customers may like or dislike.

Often, you might be overwhelmed by the number of variables you have to manage and analyze all at once. Although you may have a highly skilled data analyst at hand, it's still time-consuming and labour intensive to manually and quickly sift through the sheer volume of data to find the optimal predictive model.

To create the best predictive models of retention, rely on the power of machine learning to quickly and accurately uncover the underlying reasons why customers are churning or why they're loyal to your brand.

Machine learning uses math, statistics and probability to find connections among variables that help optimize important outcomes such as retention. These models are then applied to new customer data to make predictions.

Machine learning algorithms are iterative and learn on a continual basis. The more data they ingest, the better they get. Compared to human performance, they can deliver insights quickly thanks to the processing capability of today.

For example, you can use analytics to identify which up-sell or cross-sell products will be the most relevant based on your customer's past purchase or browsing history.

Often, companies don't have employees with high-level analytics (data science) skills. Third party providers can provide a solution that automates data integration and analysis.

4. Get data-driven insights with text analytics

To get deep, data-driven insights, don't forget to analyze your free-text responses to your open-ended survey questions. If you don't you may well miss them!

You can do this with text analytics solutions. With a text analytics tool that uses sentiment analysis, it's easy to spot customer pain points.

And, if you collect lots of data, make sure you actually use it. One study found that only 15% of senior leaders actually use customer data consistently to inform business decisions (Harvard Business Review).

At Thematic, we have developed an AI algorithm that automates analysing free-text feedback in surveys using machine learning and natural language processing, and in essence, simplified the way businesses are getting insight from their customer data.

5. Segment to focus on retaining the right customers

Using data analytics to segment people into different groups, means you can identify how each segment engages with your brand and product. This then allows you to look at each subgroup and draw insights, followed by adopting different communication and servicing strategies to increase retention for your most wanted customers.

Analyze data such as your customer demographics, lifestyle, products purchased by each category and type of customer, the frequency of purchase and purchase value. In this way, you'll discover which type of customers are driving the most revenue. Some cost too much to deliver revenue, so you'll know if you want to focus your efforts on.

Understanding the difference between these types of customers, can in some cases make or break a business, especially if you're just starting out. Knowing customer value is crucial to be able to make critical decisions. You can segment by historical value, lifetime value, value over the next year or the average customer value by segment. Using the right segmentation, you'll then create highly targeted product recommendation offers. Segment your customers to offer relevant discounts for different channels (in-store, online, mobile). Mix it up a bit, every customer doesn't have to receive the same offer.

Another useful way to use segmentation is to monitor the time-sensitivity and seasonality of your promotional codes. By monitoring sales data, you can see whether these codes are redeemed more often in the morning or afternoons or perhaps straight after a sales communication. The more you know about what a demographic responds to, the more you can focus on taking the right actions.

CUSTOMER ANALYTICS CUSTOMER SEGMENTATION



While the service needs of the customers remain individual, it is more and more difficult to approach customers individually through centralised operations. Customer segmentation identifies groups of customers that are similar such that when you apply different service approaches to limited number of customer segments; it is the optimum compromise between the benefits of the centralised operations with the individual approach to each customer.

Customer Segmentation and Profiling Analysis identifies customer segments with similar traits, attributes and characteristics that help marketers use smart analytics methodologies to target these customer segments with tailor made offerings (product, price, promotion and channel) and enhance customer acquisition, life time value , retention and loyalty through effective marketing and churn prevention strategies.

India-specific challenges

Targeted marketing can help Indian brands get ahead in the competitive environments they operate in, provided they surmount the following challenges:

Balancing product marketing with targeted marketing: India's best marketers are trained in fast-moving consumer goods (FMCG), and are very effective product marketers, focusing on brand building, product awareness and market share growth. When balanced with targeted marketing, product marketers get the support they need to pull the customer from awareness to retention and ultimately, advocacy.

Limited locational data: Indian brands currently have limited access to data that informs them about their customers' locations. However, with the explosion in the use of mobile devices, we can expect this to change in the near future. This will open up the option of using locational data to

target a customer as she walks into a store or even into just the vicinity of a store. The only snag might be data privacy concerns, which are the biggest threat to location marketing in the West.

Limited customer behavioural data: Many companies in India boast of huge customer databases, but more than 50 percent of names in those databases are either inactive or not accurate. Furthermore, the data that is accurate is not particularly predictive from the perspective of targeted marketing because it lacks detailed information on behaviour. Brands must create strategies to pick up genuinely useful data on what individual customers are buying, when and where they are buying and what they are redeeming over time.

CROSS- SELL OR UP SELL MODELS

Cross-sell involves the sale of multiple products offered by a single product/service provider to a new or existing customer. Up-sell is selling higher value products/services to an existing customer. When approaching customers during the customer lifecycle, the key consideration is customer value for the company. For high value customers it is worth spending more time and resources in servicing, marketing and support. The most valuable customers are the ones that are buying new products, actively using them and staying loyal to the organisation. Cross sell and upsell are ways to increase customer value. The magic of successful cross sell is to make a relevant offer to the right customer at the right time and using the right communication channel.

For example:

1. You plan to purchase a mobile phone within a price range of Rs. 30,000(~\$500). However, you eventually end up purchasing a mobile phone of Rs.42,000 (\$650) because the salesman presented various other phones with fantastic features and you got swayed away with them. (This is Up-Selling).
2. You plan to purchase a mobile phone worth Rs. 30,000(~\$500), but the salesman offered you a charming deal of buying mobile phone with exclusive JBL headphones for Rs.40,000 (~\$634) only and you again got swayed away. (This is Cross-Selling).

Cross-selling is a core component of a customer centric relationship strategy and requires an integrated view of the customer. The success of a cross-sell program depends on enablers such as organizational commitment; well-defined business strategy; effective execution; regular monitoring; and effective targeting strategy. Cross-selling has proved to be a defining strategy for profitable growth across multiple sectors.

Benefits of Cross Selling

Cross Selling offers benefit to both the ends of marketing cycle i.e. customer and firm.

For the Firm

- Builds customer equity

- Differentiates from competition, enhances market position
- Promotes diversification and innovation
- Stimulates universe expansion and entry into new markets
- Balances growth between new and existing customers, low and high margin products and segments
- Enhance customer profitability
- Discourages customer attrition, improves customer loyalty

For the Customer

- Patronizes the brand
- Broadens choices of product and services
- Offers convenience through one-stop shopping, flexibility, consolidated bill and others
- Increases customer satisfaction
- Lowers price
- Encourages better customer service from relationship marketing

Industry Exemplars

- **E-commerce:** The long term strength of Amazon is its ability to recognize customers one-on-one and effectively cross-sell across categories based on targeted recommendations. The decision engine is “collaborative filtering”, which recommends items deemed to be similar to the items that the “user or a similar like minded consumer” liked in the past based on expanded view of customers’ purchase history and behaviours.
- **Retail:** Segmentation based cross-sell and product diversification is the key enabler for Tesco to reach top supermarket status in Britain. Targeting is based on differentiated profiles using segmentation and effective data usage of loyalty program, demographics and lifestyle attributes.
- **Financial Services:** Cross-sell is a core strategy for revenue growth for Wells Fargo, which has the highest cross-sell ratio in the industry at 6 products per household. The bank is an adopter of a successful bundling strategy backed by strong analytics support and effective tracking of cross-sell initiatives.

Cross-sell Offer Strategy

Bundling the sale of products or services together as a combined offering using demand for the primary product to sell the secondary product. An example is a package / combo deal. This combined product or offer is at a discount so that it is more attractive to buy the bundle than products standalone. For example – Selling a combo meal for McDonalds is a good example of bundled sale.

Sequential cross-sell involves selling different products or services at different points of the customer's tenure with the firm. A good example of this is life-stage marketing, i.e. if a customer has bought a printer from you 6 months back, it might be time to check if he needs a cartridge. Varying strategies tailored to different segment level profiles constitute an integral component of a customer centric approach to cross-selling.

What is down-selling?

Although online stores often look for ways to sell more expensive products and add-ons, offering customers less-expensive items can also increase profits in many cases. That's where down-selling comes into play.

Down-selling is the opposite of up-selling. It can be effectively used when a customer is trying to back out of a purchase. At this point, you need to adapt your offer to the customer's budget and provide a better (that is, cheaper) price for another item that has similar features to the original item. This approach will give you a better chance to be accepted, and selling something is always better than nothing.

Here's another example of mobile phones. If a customer can't afford a premium mobile phone, you can recommend a less-expensive alternative, such as last year's model.

Here's a graphic of the mobile phone examples to help illustrate the difference between upselling, cross-selling, and down-selling:



UNIT VI: APPLICATIONS OF HR ANALYTICS

The technological advancements like AI, ML and Data Analytics have caused disruptions in almost all industries around the globe including the recruitment industry. In fact, it can be said that Data Analytics has revolutionised the entire process of retention and recruitment.

Analytics has facilitated HR professionals with numerous abilities - evaluating the recruiting process and conversion rates, exploring the areas of improvement and identifying the best fit for a job. The role and significance of data in recruitment are expected to enhance in the coming years as the preferences of hiring companies keep evolving and attracting the right candidates becomes one of the key challenges HR leaders face.

How important is Data to Recruiters?

Every HR manager is constantly struggling with some general issues - getting best fit for a job role, saving time for HR personnel involved in the hiring process, building more diverse teams and avoid bias, measuring outcomes of employees, avoiding guesswork in hiring process etc. Recruitment analytics play a key role in addressing these challenges, tapping into the pool of best candidates, save time, make better business decisions and maximise employee productivity.

Let's dig deep into how recruitment analytics can help the hiring managers to brush up the hiring process.

- **Enhance the Quality of Hiring:** With increasing competition and requirement for skilled candidates for evolving job roles, organisations are having a difficult time finding the right fit for their job openings. Recruiters are seeking better quality tools and technology to boost their performance as it is quite difficult to find the best fit for various openings just by screening through numerous resumes. Various technological tools like a resume parser are becoming popular among hiring teams across organisations as it does the work in a few clicks what otherwise takes many days - extracting data from resumes and putting it into pre-defined fields.
- **Enhance Your Company's Image:** Data can deliver benefits beyond one's imagination. In addition to helping in the recruitment process, it also helps in enhancing employer's brand value. The hiring teams can conduct sentiment analysis or ask the applicants to fill out a survey to gauge how the candidates see their company. The responses will help in knowing if your brand value is positive and the areas of improvement.
- **Ensure Diversity in the Workforce:** Diversity across teams and within a team is important in today's ever-evolving corporate world. HR managers can rely on data to bring the right people to the organisation and even track their diversity initiatives. For example, if an organisation's objective is to increase the percentage of female employees in their workforce by 25%, the hiring managers can choose the metrics in the hiring process accordingly.
- **Find Your go-to Job Platforms:** The hiring teams of every company, big or small, use certain job platforms while searching for the perfect fit for a job role. While some of

these job boards deliver substantial results, some give poor ROI. With analytics, HR professionals can analyse the usefulness of these platforms to their organisations.

- **Building Future Hiring Plans:** In addition to using data for your current hiring needs, it can also be used to build your future hiring plans or predicting hiring needs for the upcoming period. With data analytics, you can see which teams need to expand or will need additional talent soon. Moreover, predictive analysis helps in predicting attrition too and many large corporations have already invested heavily in predictive analytics in their attrition and hiring processes.
- **Data Driven Policies:** When you work on the basis of data instead of regular HR tracking, it ensures that you are working on evidence based information. It also helps when you have to demonstrate the value of new hiring policies to your team - present statistics and facts to back up your efforts.

RECRUITMENT ANALYTICS

Recruitment analytics is a combination of data and predictive analysis that provides real-time information to help you hire faster. Powerful recruiting analytics helps you explore every aspect of your business, turn data into actionable insights, and make better recruiting decisions faster.

Every great hire gives you an opportunity to broaden your network, build new relationships, and spot talent for the future. Predictive analytics in recruitment helps you to reduce cost per hire and find talents faster. Streamline your broad candidate database from records and get a head start on filling that talent pool with recruitment data analytics. When there are more open positions analytics for recruitment becomes essential in measuring KPIs.

When there is more than one open position, recruitment analytics become essential because they offer a high level view of your whole hiring effort. And with it answers to these questions:

1. Where are your hiring bottlenecks?
2. Which hiring managers need help?
3. Which positions need urgent attention?
4. Which are your best sources for hires?

Too much of recruiting analytics has been about calculating the cost per hire. Cost per hire is calculated by adding up all of your recruitment costs from ads to external recruiters, referral bonuses, plus your own hiring team's compensation and benefits costs, and dividing it by the total number of new hires for the calendar year.

As well as being tough to meaningfully calculate, for smart companies it may be the wrong place to look. The point is not to hire more cheaply, it's to get better results from hiring.

Resume screening: Using tech to be a better gatekeeper

Technology can reduce the time and cost of initial candidate screenings, as well as human bias.

Where humans once suffered paper cuts from reading stacks of physical resumes, algorithms now slice and dice as many applicants as a company needs to review. If companies are expanding their talent pools with more professional networking sites, it's critical to sort through the noise by filtering search results and candidate pools by desired experiences, skills, selling results and qualifications.

One can question whether it's wise to trust an algorithm to make decisions about who to let through a company's gates. After all, wouldn't humans have a better read on the intangible qualities of a top performer? And can a machine deliver on recruiting standards while increasing employee diversity?

It turns out that algorithms may do a better job than humans when sorting candidates, both from a performance perspective and from a bias perspective. Companies that embrace automated resume screening systems outperform their human review processes by at least 25%, resulting in better outcomes and more diversity. This finding comes from researchers from the University of Minnesota who reviewed 17 studies of companies and academic institutions that backtested their applicant evaluation algorithms and compared them to their traditional methods.

Here's a real-world example of this from a professional services company that struggled keep up with screening the 250,000 job applications it received per year.

To lighten the load, they introduced a resume screening algorithm that mined for education and work experience and considered the history of the company's past applicants, extended offers, and accepted offers. The tool sorted resumes into 3 groups: most likely to be hired, least likely to be hired, and in-between applicants that needed the attention of a human recruiter. The intent was to reduce time and cost for the HR team, but there were concerns that a machine would interfere with the company's goals to hire more women.

The algorithm delivered a 500% return on investment in terms of cost savings.

It sorted away 55% of the resumes, either into the "least likely to be hired" pool (where they were automatically rejected), and the "most likely to be hired" pool (where they automatically went on to the next stage).

Moreover, it improved the female hiring initiative. The system passed 15% more women than manual screenings did, and all of them on merit. "The foundational assumption—that screening conducted by humans would increase gender diversity more effectively—was proved incorrect."

Google also uses a systematic, data-driven approach to reduce bias in resume screening. They analyze employee profiles that lead to success at Google, and use the insights to reassess candidates they are about to reject. This helps them identify potential false negatives, helping them catch excellent hires they would otherwise pass over.

In a similar exercise, an Asian bank challenged their hiring assumptions by taking a closer look at the backgrounds of their top performers:

“Whereas the bank had always thought top talent came from top academic programs, for example, hard analysis revealed that the most effective employees came from a wider variety of institutions, including five specific universities and an additional three certification programs. An observable correlation was evident between certain employees who were regarded as “top performers” and those who had worked in previous roles, indicating that specific positions could serve as feeders for future highfliers.

Both of these findings have since been applied in how the bank recruits, measures performance, and matches people to roles.

Impact category	Mechanism	Example platforms
Recruiting and talent acquisition	Find better candidates	<ul style="list-style-type: none">• LinkedIn• Monster• SmashFly• ZipRecruiter
	Discover hard-to-find, niche talent	<ul style="list-style-type: none">• Dice• Entelo• Hired• LinkedIn• Niche Talent
	Access non-traditional workers or channels	<ul style="list-style-type: none">• LinkedIn• TalentBin• Tomigo
	More efficiently filter to select interviewees	<ul style="list-style-type: none">• Chequed• ClearFit• Hire IQ• TalentWise• TrueAbility
	Use candidate data for better assessment	<ul style="list-style-type: none">• Codility• JobFig• Pomello
	Tailor approach to each candidate	<ul style="list-style-type: none">• Future providers to emerge

Source: McKinsey



The results: a 26 percent increase in branch productivity (as measured by the number of full-time employees needed to support revenue) and a rate of conversion of new recruits 80 percent higher than before the changes were put in place. During the same period, net income also rose by 14 percent.

COMPENSATION ANALYTICS

Compensation is one of the key culture-definers for organizations. Depending on how [compensation is] designed, communicated, and managed, it can positively or negatively influence an organization's culture and impact an organization's optimal performance. Compensation management is the practice of ensuring that an organization is paying employees fairly and competitively. Compensation managers work with employment data to understand all of the factors that go into pay and whether or not the benefits an organization offers meets the needs of their workforce.

The objective is to make informed compensation decisions to attract and retain the right talent. This is why it's so important to base these decisions on relevant and accurate data. Organizations need to analyze compensation programs to figure out what's working and what isn't.

A compensation analysis uses internal and external data to determine whether an employer rewards employees fairly or not for the work they are doing.

Here are a few key concepts in compensation analysis that you need to understand:

External competitiveness – Employers compare their compensation data and practices to external companies. For example, Salesforce found that their engineer salaries were comparable to engineers at Microsoft. It was a fair comparison because of the size and geographical spread of both organizations. On the other hand, engineers at Slack were paid much less, but they are a much smaller organization.

Internal equity – Employers compare employees' salary and indirect compensation data to ensure fair compensation for the level and type of work done.

Region – Employers compare compensation data of people doing similar work within a particular region.

Level – Employers compare employees' levels and the level at which they are compensated.

In all of the components mentioned above, it's noteworthy to mention that even though salary data is important to look at, it is only one part of compensation. There are other employee benefits such as medical care, discounts, car, share schemes, and housing allowances.

A key characteristic of compensation is that it is dynamic. This means that it continually changes and progresses based on the internal and external environment.

Benefits of conducting a compensation analysis

A compensation analysis is a vital component of an organization's talent management strategy, as it helps attract and retain the best employees on the market. Let's take a look at other benefits of a compensation analysis:

- **Salary benchmarking** gives an impartial idea of competitive salaries and allows organizations to make informed decisions. Salary benchmarks provide data points, whether it is worth it or not to pay an employee above the average salary. It also helps understand the holistic remuneration packages offered by employers.
- **Evaluating pay equity** allows organizations to compensate employees doing the same level of work in a fair way. Conducting a comprehensive compensation analysis also enables you to correct historic pay gaps.
- **Transparent compensation** decisions leave the decision-making of salaries in the hands of accurate and impartial data. This leaves employees with a higher level of trust in the organization and their managers.
- **Compensation analysis is dynamic**, as mentioned before, and thus, you're able to make projections based on future needs or employees and how this may affect your compensation strategy.
- **Identifying opportunities** – Through a comprehensive compensation analysis, you can identify where you can improve your compensation strategy. You may be able to find different ways to remunerate employees, for example.

TALENT ANALYTICS

Talent analytics is the systematic analysis of workforce data. It's data that's collected and used by HR teams to help you better understand your employees and your potential hiring pool so you can uncover any opportunities or gaps that need to be addressed.

Laszlo Bock, former Senior Vice President of People Operations at Google, nicely sums up the value of talent analytics. "By analyzing behaviors, attitudes, personality traits, and perception over time, we aim to identify the biggest influencers of a satisfying and productive work experience."

By collecting and analyzing employee data, "[It] allows us to flex our people practices in anticipation of our peoples' needs," said Bock.

Talent analytics is also known by many other names, including human resource analytics, workforce analytics, human capital analytics, people analytics, or hiring analytics.

Examples of talent analytics in action

In practice, talent analytics can be used for many different HR purposes, but always includes an infrastructure of collecting and reporting on data about your employees. In most cases, this includes elements like a reporting dashboard with key HR metrics, and regular employee surveys or tests.

For example, all of Accenture, Intel, IBM, and Twitter use sentiment analysis from surveys or internal company chat messages to evaluate how their employees are feeling about the company. They then use this data to identify problems or opportunities that didn't come up through typical review processes.

Wal-Mart and Credit Suisse Bank, on the other hand, feed a massive amount of data points into an advanced algorithm to predict which employees are most likely to quit. This helps them forecast where they need to focus their hiring efforts before the positions even need to be filled.

Other companies use talent analytics to create internal platforms where they can match current employees to vacant role opportunities within the organization based on their skill profiles.

Another use case would be using talent analytics to evaluate the characteristics and skills of your company's top-performing salespeople. You could then test candidates on those skills as part of a pre-employment assessment.

An untapped opportunity

The importance of talent analytics is clearly understood by HR teams, which are already adapting to this new world: 30% of corporate HR departments already have at least one team member or an entire team dedicated full time to analytics.

LinkedIn's Global Recruiting Trends report from 2018 found similar results. According to their research, 50% of hiring professionals were using big data as part of their strategy.

And yet, most HR teams know they aren't taking full advantage of data yet. According to Gartner, "most companies are not realizing the value from their analytics investments, with only 21% of HR leaders believing their organizations are effective at using talent data to shape talent acquisition and recruiting strategies, improve employee engagement and inform other business decisions."

And there's one area of HR that stands to benefit the most from talent analytics: hiring.

How talent analytics can help with hiring

Hiring the wrong people is expensive. According to a CareerBuilder survey, companies lose \$14,900 on average on each bad hire—with 74% of companies surveyed claiming to have made at least one bad hire.

According to the same survey, the most common reasons for bad hires were:

- The hiring manager assumed an under-qualified candidate could learn the needed skills
- The candidate lied about their qualifications
- The hiring manager took a chance on a nice person

- The hiring manager felt pressured to fill the role quickly
- There weren't any other qualified candidates

All of these problems could have been avoided if talent analytics and HR data were used more proactively.

Because with talent analytics, you can put more data and strategy into the hiring process rather than just focusing on filling the position.

Here's how talent analytics can benefit the hiring process.

1. Identify skills on your current team

One simple way to hire better candidates is to fill skill gaps missing on your current team.

Current employees can be tested on their hard and soft skills to determine where your team may be lacking. You can then leverage that data to hire employees that make a bigger impact.

Alternatively, you can use skill assessments to identify diamond-in-the-rough employees who may be able to excel in a bigger role.

2. Identify representational gaps & improve diversity

There are countless benefits to diversity in the workplace.

Diverse teams are more innovative and creative. They are faster problem solvers. They make better decisions. They are more productive and perform better. The list goes on.

By using talent analytics to monitor the diversity on your team (whether it's race, gender, or any other measure of diversity), you can then use this data to inform your diversity hiring initiatives.

3. Uncover key recruiting metrics & improve the candidate experience

By tracking key recruiting metrics like application completion rate, applicant-to-interview rate, interview-to-offer rate, offer-acceptance rate, source of hire, quality of hire, and turnover rate, you can unlock new insights about your hiring process.

By tracking and learning about your strengths and weaknesses at each stage of the hiring process, you can uncover any weak points in the candidate experience.

For example, you may uncover a flaw in the application process or learn where you need quick response times or more candidate touchpoints.

This will help you optimize your hiring process, ultimately reducing your time to hire and your recruitment cost per hire.

4. Hire with less bias

By making data-driven hiring decisions, you can work to minimize unconscious bias from your hiring process.

And this goes way beyond classic examples of hiring bias, such as interviewer bias. You may discover something totally unexpected about what makes a great hire.

For example, AT&T and Google discovered through talent analytics that having a stellar academic record at a prestigious school was a less predictive measure of job success than having a demonstrated ability to take initiative. Without using hiring data, they never would have eliminated this non-predictive hiring bias.

5. Better predict candidate success & hire better candidates

All of the above leads to a better understanding of what makes a great candidate.

The aforementioned LinkedIn report found that companies who implemented talent analytics as part of their hiring process reported that it helped them with talent acquisition—increasing their employee retention by an average of 56%.

6. Improve forecasting

While your forecasting model may not be as complex as Walmart's, you can use talent analytics to track historical hiring data, employee turnover, internal job changes, and factors like the current hiring rate.

Armed with this data, you can improve succession planning and predict any potential job openings or gaps in your team before they actually happen—giving you more time and less pressure to fill a role when it opens up.

The many benefits of using talent analytics are obvious. But how can you actually introduce talent analytics into your hiring process?

How to use talent analytics in your hiring process

If you're new to talent analytics, this can all seem pretty daunting to put into practice.

But fear not, there are some realistic ways for you to start using talent analytics to boost your hiring today.

1. Start with a data-driven mindset

Some parts of HR, like evaluating candidates, are all-too-often, gut-driven. Other parts of HR, like employee reviews, can feel too process heavy.

Committing to a data-driven mindset can help alleviate both issues. Gut-driven processes can be informed with data, and process-heavy initiatives can be lightened with easier-to-interpret data.

The key is to commit to a culture of using data on your HR team.

This starts with ensuring that everyone on your HR team has basic data analysis skills and that you encourage your team to back up their decisions with data rather than intuition.

For example, Google once posted an HR business partner job description where they preferred candidates with “strong analytical and problem-solving skills with proven ability to organize and analyze data, using HRIS systems for reporting.”

Facebook has similar requirements, requiring HR Business partners to “drive data-led decision-making through analysis of key people metrics.”

2. Use an end-to-end HR system

Unfortunately, you won’t be able to make data-driven decisions unless you have a system that gathers data. To run a successful talent analytics program, you will need a true end-to-end HR system or technology stack that allows you to track employee data at all stages of their lifecycle, starting from the application process.

That’s why the IT team should be your best friends. Partner with your company’s IT team to develop data governance principles and to ensure your hiring team is equipped with the right HR technology. In fact, many HR teams now have an IT specialist working directly for their team.

The key to this is implementing a system that makes talent analytics user-friendly for your team. If data isn’t easy to access and easy to interpret, it will never be used.

3. Identify specific challenges and KPIs

Data for the sake of data doesn’t get you anywhere. But once you commit to a data-driven culture, you can more easily identify the problems you want to solve as a hiring team, and then track your progress towards those goals.

Identify specific challenges you want to address and then identify the key performance indicators (KPI) that will help you measure your progress.

For example, if you commonly face the issue of having many vacant roles, you would likely want to track time to hire and turnover.

4. Build your ideal candidate persona

Based on what you learned from data from your current employees, identify any skill or representational gaps on your team.

Use this information to build your ideal candidate persona. Ask questions like:

- What role-specific skills do they need?
- What values should they possess?
- What language skills do they need?

You should also assess their long-term motivations, like whether or not you want to bring in someone who could eventually move into a leadership role, or whether you'd prefer to hire someone who's comfortable staying in the role long-term.

5. Revamp & track your employer brand

Use your ideal candidate persona(s) to reassess your employer branding.

Make sure your website, social media profiles, and job descriptions appeal to the applicants you're trying to target. For example, if you're hiring for more diversity, make sure your job descriptions use inclusive language.

Then, as discussed above, track your recruiting metrics like inbound traffic to your job postings, application rate, your offer-to-hire rate, and other metrics that may help you learn how to improve the candidate experience.

6. Consider screening candidate resumes

Resume screening tools have become a popular way for large companies to automatically sift through the thousands of resumes they receive on a daily basis.

Resume screening software is powered by AI that filters applications based on certain keywords, saving talent acquisition teams lots of time.

However, resume screening tools aren't without their issues. Applicants are quickly learning how to include target keywords in their resumes, leading to false positives. On the other hand, some qualified candidates may be erroneously filtered by the system if they use non-standard phrasing or resume formats.

7. Use pre-employment assessments

A better alternative to screening candidate resumes is to use pre-employment assessments.

A pre-employment assessment is a package of tests given to a candidate to complete as part of the application process.

Potential test types include:

- Cognitive ability tests
- Language tests
- Personality & culture fit tests
- Situational judgment
- Programming skills
- Software skills
- Other role-specific skills

By using a combination of tests, you get a clear, data-driven view of each applicant. Not only does this save you from sifting through every resume, but it also gives you a less biased perspective on an applicant's true value and skillset.

Take a data-driven approach to hiring

To make talent analytics part of your hiring process, make TestGorilla part of your HR tech stack.

You can easily create high-quality, customizable assessments for any job role, allowing you to choose the tests and questions that work best for you.

You can then analyze assessment results in real-time from a convenient candidate dashboard, allowing you to instantly compare your candidates by the metrics you care most about.

TRAINING ANALYTICS

Learning and development teams spend many hours developing training courses for employees. But without a sufficient feedback loop, questions are raised: is this training effective? Do employees find it useful? Are we properly supporting organizational goals and improving performance? Without properly set up training analytics, it can be anyone's guess as to whether the carefully curated training program is effective.

Training needs analysis is a process that a business goes through in order to determine all the training that needs to be completed in a certain period to allow their team to complete their job as effectively as possible, as well as progress and grow.

There are 3 key steps involved in training needs analysis to ensure your business is making the most of the process:

1. Decide On Skill Sets

The first stage is to decide on the skill sets that you require all your team members to have in order to do their jobs properly. This means looking at every job role within your business separately and considering things like the different departments or levels of seniority which will affect this as well.

2. Evaluate The Skills Of Staff

The second stage is to look at all your team members and evaluate their current skill levels in relation to the skills you have laid out in the first stage of this process. This will allow you to see who is meeting your expectations, and who needs to complete further training in order to meet the expected skill level.

3. Highlight The Skills Gap

Now that you know where you want your team to be and the level they are currently at, you will easily be able to see the gap (if any) that has appeared between the two. Now you know what the gap is, you need to use training to help close that gap and ensure your team is at the level you expect them to be.

Benefits For Your Business

1. Identify Knowledge Gaps Before They Become A Problem

One huge benefit of conducting training needs analysis is the fact it can help you identify any knowledge gaps your employees may have before it becomes an issue. It's better to highlight a potential problem and tackle it head-on, rather than becoming aware of the skills gap when an issue arises because of it.

The training needs analysis will allow you to take a proactive approach rather than waiting for something to go wrong before you realise there is a problem.

2. Helps You To Plan Your Training For The Year

Another huge benefit of training needs analysis is that it makes it much easier for you to plan your training for the upcoming year (or whatever block of time you work with). Once you have identified the skills gaps that exist in your business, and then all the staff members who need additional training in certain areas, it's easy to pull together a training plan which will cover all these skills gaps.

Rather than trying to guess the type of training that will be most useful to your organisation, or who needs to complete the training, your training needs analysis will make the whole task much easier, and you can be confident that the training you have selected will make a direct impact on your business!

3. Highlights Training You May Not Have Considered

It can be hard to sit down and plan out a training schedule for a large organisation without completing some sort of background research first. You may think that you know the type of training your team should be completing, but training needs analysis could actually highlight a whole load of areas that your team needs training on that you never even considered before.

That's why training needs analysis is so useful because it can highlight training needs you may not have considered before and show that you need to start offering training in different areas to ensure your staff are performing at their best.

Without the use of training needs analysis, you may never have considered a particular area of training, which could have severely hindered your business.

4. Ensures Your Training Is Focussing On The Right Areas

As we said above, it's important to have concrete reasons for adding training to your training schedule, as you can't just assume what is and isn't important for your team to learn. Completing a training needs analysis will allow you to see exactly what you need to focus on, but it will also highlight the areas your team really don't need any further training on for the moment.

If there are no apparent gaps in knowledge in a particular area, then running further training on it could be a waste of time and money!

5. Helps To Decide Who Should Attend Which Training Sessions

Another important step in planning training is to ensure the right people are in the right training sessions. There is no point in making everyone in your organisation attend every training session you run. It's a massive waste of time and money for your business, and staff won't be engaged with training sessions if they are frequently attending training which is of no use to them.

A training needs analysis will enable you to target the correct people for each training session, ensuring everyone is following a personalised training plan, so they get the most benefit possible.

6. Helps You To Prioritise Training Needs

When it comes to planning out your training, it can be hard to decide which training sessions are the most important. However, training needs analysis can help you pinpoint the training which needs to be completed ASAP, and which training can be left till later down the line.

When you think about the skills that each team member needs to have, you may want to prioritise these regarding how key they are. For example, if you have a customer-facing team, ensuring they have top-notch customer relations skills may be top of the list.

If you notice a gap in the face-to-face skills for some of these employees, it only makes sense that you would want to tackle this first, as this is a key aspect of their job role, and lack of training in this area could have a negative effect on your business.

All other training can be prioritised afterwards, but it's important to get that customer relations training booked in and attended as soon as possible to make sure your customer-facing teams are top performers.

HUMAN RESOURCE RETENTION ANALYTICS

Modern business relies on positive employee retention. When companies experience high employee turnover, the costs of hiring and training create financial stress. When vital positions remain unfilled for extended periods, current staff become taxed and productivity suffers.

Escalating employee turnover is why it is essential that companies integrate intelligent, intuitive human resource protocols that target the problem. When a company can predict turnover, it can take steps to prevent it.

Many companies leverage internal employee survey data as a way to compile data-based insights into employee welfare. Are employees happy? Just ask them in a survey. Unfortunately, valid employee survey data relies heavily on people to be truthful. Case in point, a Harvard Business Review survey discovered that 58 percent of workers trust strangers more than their bosses. Such distrust of management can cause employee survey information to be unreliable. The situation worsens when we consider that many companies don't grasp the reasons why employees flee from positions. 89 percent of employers believe that turnover stems from an employee's desire to earn more money. However, as the study reveals, only 12 percent of employees leave their jobs over salary. That's an enormous (and costly) misperception about employee turnover.

When our information is compromised, we tend to make poor business decisions. Your company can improve its retention strategies by relying on data analytics that removes human emotion.

Don't ask the employees how happy they are. Instead, consider the variables that factor into their happiness. The good news is that you already possess important employee data.

To predict employee turnover, consider hard values not associated with a person's emotions. Instead, look for consistent factors across your enterprise, such as

- Benefits
- Promotions
- Past reviews
- Historical pay
- Sick time used

Once you've compiled data, you want to have an analyst relate it to trends. For example, does more or less sick time used correlate to employee turnover? Can an analyst confirm a lack of employee promotions instigates increased turnover?

It is important to allow the data to speak for itself. For example, it may feel intuitive to surmise that employees with more historical promotions are less apt to quit. However, if a history of promotions seems to reveal a trend of employee turnover, promotions (or increased responsibilities) is the issue that deserves further analysis.

Such data shouldn't signal immediate and drastic companywide change but rather deeper scrutiny into employee promotion logistics.

Predicting Employee Turnover Offers Warnings, Solutions

When your company predicts high-risk employee turnover demographics, you increase your enterprise's ability to address the problem. In other words, you may be able to lessen the turnover percentage by focusing on the specific types of workers likely to resign.

Additionally, your enterprise may invoke strategic changes that treat the issue from a global perspective. If your company can make a widespread change in policy that increases employees' odds for staying, it automatically reduces turnover costs.

Current data suggests that employees leave their jobs for salary reasons less often than we assume. This means your company may be able to lessen employee turnover without increasing salaries. In such cases, data used in predictive analytics can save your company money on multiple fronts: hiring expenses, training costs, and overall payroll expense.

A company leveraging survey data can use the results to complement predictive analytics. One should not interpret this article as condemnation for employee surveys, rather, a deep dive into the aptitude of in-house analytics.

Predictive Analytics Can Improve Culture, Productivity, and Achievements

Companies rightfully focus on employee retention numbers, but data analytics might also help improve a company's culture and productivity. For example, if a company concentrates on salary's influence on employee happiness, it might miss other employee-happiness variables, causing it to miss out on changes which might actually create a better work environment.

When employee predictive analytics becomes company culture, it cuts down on reactionary decisions. It increases the odds of company success by exposing the least costly obstacles to that success.

WORKFORCE ANALYTICS PROJECT WORK

Workforce analytics is the approach of measuring behaviors of people (candidates, employees) and analyzing them to improve people and business performance. This is done by analyzing people data using statistical methods and software in order to make better workforce decisions.

For example, using workforce analytics, managers would be able to:

- Make better hiring decisions by predicting candidate success
- Prevent talent from quitting their job by predicting employee turnover
- Test which employee policies are effective – and which ones are not
- Identify and quantify work accident risk
- Analyze future workforce need
- Optimize the employee experience
- Link HR actions to business outcomes

When workforce analytics is done well, it's not only the business that benefits. Candidates and employees can also benefit. They can, for example:

- Make a better fit between their wishes and needs and the organizational demands
- Develop themselves and utilize their potential better
- Use learning solutions that are tailored to their actual performance
- Get granular feedback that can be used to improve performance.
- Be enabled to contribute more to the objectives of the organization

Workforce analytics relies on up-to-date employee data, transparency, and buy-in from the employees themselves.

From people analytics to workforce analytics

Currently, the general opinion seems to be that people analytics is a better label than HR analytics. Increasingly the workforce is consisting of more than just people. Robots and chatbots are entering the workforce. The first legal discussions have started: who is responsible for the acts of the robots? If we're also analyzing robots, we're moving from people analytics towards workforce analytics. Robot wellbeing and robot productivity is a nice domain for HR to claim.

For more detailed information, visit the following links:

- <https://www.kdnuggets.com/2015/10/beginners-guide-predictive-workforce-analytics.html>
- <https://www.kaggle.com/haijie/hr-analytics-project>
- <https://www.peoplematters.in/article/hr-analytics/how-nokia-is-tackling-the-covid-19-crisis-with-analytics-26630>

CES PARAMETERS & HELPFUL INFORMATION

For Internal Circulation Only
VIMR

CASE STUDY(s)- based on Visualization and Data Issues

1. Case Study – Walmart

Walmart is the largest retailer in the world and the world's largest company by revenue, with more than 2 million employees and 20000 stores in 28 countries. It started making use of big data analytics much before the word Big Data came into the picture.

Walmart uses Data Mining to discover patterns that can be used to provide product recommendations to the user, based on which products were brought together. WalMart by applying effective Data Mining has increased its conversion rate of customers. It has been speeding along big data analysis to provide best-in-class e-commerce technologies with a motive to deliver superior customer experience. The main objective of holding big data at Walmart is to optimize the shopping experience of customers when they are in a Walmart store. Big data solutions at Walmart are developed with the intent of redesigning global websites and building innovative applications to customize the shopping experience for customers whilst increasing logistics efficiency. Hadoop and NoSQL technologies are used to provide internal customers with access to real-time data collected from different sources and centralized for effective use.



2. Case Study – Uber



Uber is the first choice for people around the world when they think of moving people and making deliveries. It uses the personal data of the user to closely monitor which features of the service are mostly used, to analyze usage patterns and to determine where the services should be more focused. Uber focuses on the supply and demand of the

services due to which the prices of the services provided changes. Therefore, one of Uber's biggest uses of data is surge pricing. For instance, if you are running late for an appointment and you book a cab in a crowded place then you must be ready to pay twice the amount.

For example, On New Year's Eve, the price for driving for one mile can go from 200 to 1000. In the short term, surge pricing affects the rate of demand, while long term use could be the key to retaining or losing customers. Machine learning algorithms are considered to determine where the demand is strong.

3. Case Study – Netflix



It is the most loved American entertainment company specializing in online on-demand streaming video for its customers. Netflix has been determined to be able to predict what exactly its customers will enjoy watching with Big Data. As such, Big Data analytics is the fuel

that fires the ‘recommendation engine’ designed to serve this purpose. More recently, Netflix started positioning itself as a content creator, not just a distribution method. Unsurprisingly, this strategy has been firmly driven by data. Netflix’s recommendation engines and new content decisions are fed by data points such as what titles customers watch, how often playback stopped, ratings are given, etc. The company’s data structure includes Hadoop, Hive and Pig with much other traditional business intelligence.

Netflix shows us that knowing exactly what customers want is easy to understand if the companies just don’t go with the assumptions and make decisions based on Big Data.

4. Case Study – eBay



A big technical challenge for eBay as a data-intensive business to exploit a system that can rapidly analyze and act on data as it arrives (streaming data). There are many rapidly evolving methods to support streaming data analysis. eBay is working with several tools including Apache Spark, Storm, Kafka. It allows the company’s data analysts to search for information tags that have been associated with the data (metadata) and make it consumable to as many people as possible with the right level of security and permissions (data governance). The company has been at the forefront of using big data solutions and actively contributes its knowledge back to the open-source community.

5. Case Study – Procter & Gamble

Procter & Gamble whose products we all use 2-3 times a day is a 179-year-old company. The genius company has recognized the potential of Big Data and put it to use in business units around the globe. P&G has put a strong emphasis on using big data to make better, smarter, real-time business decisions. The Global Business Services organization has developed tools, systems, and processes to provide managers with direct access to the latest data and advanced analytics. Therefore P&G being the oldest company, still holding a great share in the market despite having many emerging companies.



Article- based on Data Mining

The Next Data Mine Is Your Bedroom

Google wants to scan your clothing and listen to you brush your teeth.

It's a familiar feeling: Type something into Google's search bar, and then start seeing ads for it everywhere. Sometimes you don't even need to search—Google's already triangulated your desires based on your emails, your demographics, your location. Now that familiarity stands to get a lot more intimate. With a fascinating pair of new patents for smart-home technology, Google is hoping users will open their home to its trademark eavesdropping.

In the first patent, Google imagines devices that would scan and analyze the surroundings of your home, then offer you content based on what they detect. According to the patent, the smart cameras in such a device could, for example, recognize Will Smith's face on a T-shirt on the floor of a user's closet. After matching this analysis against your browser history, the device might then say aloud, "You seem to like Will Smith. His new movie is playing in a theater near you."

It doesn't stop at Will Smith movies. The patent imagines that smart-home devices would make all types of inferences about users, sorting them into categories based on what the devices see in their most personal spaces. Using object recognition, they could calculate "fashion taste" by scanning your clothing, and even estimate your income based on any "expensive mechanical and/or electronic devices" they detect. Audio signatures, too, could be used to not only identify users, but to determine gender and age based on the timbre of their voice. The smart home would recommend what to watch and where to shop, all based on how it sorts users into categories of taste, income, and interest.

If this sounds invasive, it's important to recognize that this is already happening, just online. Google and Facebook both record and analyze user behavior, use it to sort people into categories, and then target them with ads and other content. Facebook likely knows your race and religion, while Google uses your emails and search history to sort you into ad-ready brackets. Netflix infers all types of data on users based on what they watch, then serves back hyper-specific movie and TV categories. This patent simply expands the areas in which your behavior is already mined and recorded from your phone and laptop to your bedroom.

And your children's bedrooms. The second patent proposes a smart-home system that would help run the household, using sensors and cameras to restrict kids' behavior. Parents could program a device to note if it overhears "foul language" from children, scan internet usage for mature or objectionable content, or use "occupancy sensors" to determine if certain areas of the house are accessed while they're gone—for example, the liquor cabinet. The system could be set to "change a smart lighting system color to red and flash the lights" as a warning to children or even power off lights and devices if they're grounded.

While people can set goals for their children or themselves, these policies could also be “based upon certain inputs from remote vendors/facilitators/regulators/etc.,” according to the patent. That opens the door for companies to offer rewards for behaviors in the home. A household may set the internal goal of “Spend less time on electronic devices,” or “Use 5 percent less energy each month for the next three months.” Google devices could then connect to anything “smart” in the home and send you, and potentially a vendor or third party, updates on usage and screen time.

Just this month, the insurance company United Healthcare began partnering with employers to offer free Apple Watches to those who hit certain fitness goals. Insurers might also offer benefits to residents whose homes prove their fitness or brand loyalty—and punish those who don’t. Health insurers could use data from the kitchen as a proxy for eating habits, and adjust their rates accordingly. Landlords could use occupancy sensors to see who comes and goes, or watch for photo evidence of pets. Life-insurance companies could penalize smokers caught on camera. Online and in person, consumers are often asked to weigh privacy against convenience and personalization: A kickback on utilities or insurance payments may thumb the scales in Google’s favor.

For reward systems created by either users or companies to be possible, the devices would have to know what you’re doing at all times. The language of these patents makes it clear that Google is acutely aware of the powers of inference it has already, even without cameras, by augmenting speakers to recognize the noises you make as you move around the house. The auditory inferences are startling: Google’s smart-home system can infer “if a household member is working” from “an audio signature of keyboard clicking, a desk chair moving, and/or papers shuffling.” Google can make inferences on your mood based on whether it hears raised voices or crying, on when you’re in the kitchen based on the sound of the fridge door opening, on your dental hygiene based on “the sounds and/or images of teeth brushing.”

Of course, patents aren’t products, but they do represent an important shift. For a long time, the foundational metaphor of surveillance studies has been the panopticon—unending, inescapable, unwanted surveillance. Now these patents seem to hint that the age of hyper-personalization will make people willing, enthusiastic participants in the panopticon, both as subjects and as architects.

SAMPLE QUESTION PAPER

For Internal Circulation & BVIMR only

(Internal Exam-Moodle Based)

Goal of HR analytics is

- A. Transforming data
- B. Gain insights from past data
- C. Understand future
- D. Find patterns in data

ANSWER: B

Which sector has highest analytics size?

- A. Ecommerce
- B. Finance & banking
- C. Marketing & advertisement
- D. Healthcare

ANSWER: B

Business Intelligence (BI) systems do not obtain their data by which of the following means?

- A. Read and process data from an operational database
- B. Process transactions
- C. Process extracts from operational databases
- D. Process data purchased from data vendors

ANSWER: B

Business analytics can be used for which of the following?

- A. Finding patterns
- B. Verifying previous decisions
- C. Predicting results
- D. All of these are correct

ANSWER: D

The first step in the business analytic process is?

- A. Determine the needs of the business
- B. Collect the data
- C. Analyze the data
- D. Predict the outcome

ANSWER: A

This is the processing of data about customers and their relationship with the enterprise in order to improve the enterprise's future sales and service and lower cost.

- A. clickstream analysis
- B. database marketing
- C. customer relationship management
- D. CRM analytics

ANSWER: D

This is a systematic approach to the gathering, consolidation, and processing of consumer data (both for customers and potential customers) that is maintained in a company's databases

- A. database marketing
- B. marketing encyclopedia
- C. application integration
- D. service oriented integration

ANSWER: A

This is the processing of data about customers and their relationship with the enterprise in order to improve the enterprise's future sales and service and lower cost.

- A. clickstream analysis
- B. database marketing
- C. B2C
- D. CRM analytics

ANSWER: D

In data mining, this is a technique used to predict future behavior and anticipate the consequences of change.

- a. disaster recovery
- b. predictive modeling
- c. phase change
- d. Digital Silhouettes

ANSWER: B

The has always relied on statistic to determine the insurance rates. Risk-based models form the basis for calculators that are used to calculate insurance premiums.

- A. Insurance Industry
- B. Travel and Tourism Industry
- C. Agriculture Industry
- D. Health Industry

ANSWER: A

Success in the is all about having the right information at the right time. By using big data and predictive analytics, algorithms help the industry in collecting the data from a variety of data sources and support from trading decisions to predicting default rates and risk management.

- A. Telecom Industry
- B. Travel and Tourism Industry
- C. Insurance Industry
- D. Finance Industry

ANSWER: D

Predictive analytics is widely used by both conventional retail stores as well as e-commerce firms for analyzing their historical data and building models for customer engagement, supply chain optimization, price optimization, and space optimization and assortment planning.

- A. Retail Industry
- B. Telecom Industry
- C. Health Industry
- D. Finance Industry

ANSWER: A

The is also using big data analytics for enhancing customer experiences and offer customized recommendations. These firms use demographic statistics, average time spent by users on certain travel-related web pages, personal historic travel preferences, etc.

- A. Health Industry
- B. Travel and Tourism Industry
- C. Insurance Industry
- D. Telecom Industry

ANSWER: B

Another business challenge that's leading to an increase in companies relying on business analytics to drive their strategy is that customers are becoming more fickle, and loyalty to products and services is rarer than ever before

- A. Competition
- B. Animals
- C. Environment
- D. Customers

ANSWER: D

Part of addressing competitive threats is to monitor and stay one step ahead of your competition—tracking, analyzing, and integrating everything you know about your competitors into the analytics of your own company. This statement refers to-

- A. Environment
- B. Competition
- C. Business
- D. Customers

ANSWER: B

Sometimes a decision maker has made a particular decision often enough to have internalized the process of gathering and analyzing data.

- A. When There's No Data
- B. When There's No Precedent
- C. When the Decision Maker Has Considerable Experience
- D. When the Variables Can't Be Measured

ANSWER: C

The scientific process of transforming data into insights for the purpose of making better decisions refers to

- A. Data Warehousing**
- B. Business Intelligence**
- C. Cloud Computing**
- D. Business Analytics**

ANSWER: D

What is data?

- A. Information**
- B. facts and statistics collected together for reference or analysis**
- C. processed collection of content**
- D. All of the given options**

ANSWER: B

What statement below best describes why we do data analytics in business?

- A. Analytics improve our understanding of how the business works**
- B. We must show a return on the investment we make in data & analytical resources**
- C. We need specific insights to make business decisions**
- D. We have to calculate & report financial results to owners / shareholders**

ANSWER: C

Business Analysis and Business Analytics are the terms that can be used synonymously

- A. True**
- B. False**

ANSWER: B

Business Analytics can aid in efficient decision making because

- A. various tools are used**
- B. a proper evaluation is done with required analysis**
- C. it is emerging**
- D. Managers like to use new terminology**

ANSWER: B

Business analytics results in which of these?

- A. Evidence Based Decisions**
- B. Data Driven Decisions**
- C. Better Decisions**
- D. All of the options are correct**

ANSWER: D

The first step in the business analytic process is

- A. Determine the needs of the business**
- B. Collect the data**
- C. Predict the outcome**
- D. Analyze the data**

ANSWER: A

Which is NOT one of the applications of BA

- A. Financial Management
- B. E-commerce
- C. Books
- D. Social Media

ANSWER: C

Which type of Analytics refer to future outcomes?

- A. Predictive
- B. Prescriptive
- C. Diagnostics
- D. Descriptive

ANSWER: A

Data Analytics is used to

- A. Confuse Managers in achieving their Goals
- B. Help Managers to gain insights of the data
- C. Help Managers in becoming a coder
- D. None of the given option

ANSWER: B

Data can be generated by

- A. Customers
- B. Industries
- C. Businesses
- D. All of the given option

ANSWER: D

Which of the following is NOT TRUE about talent acquisition?

- A. Identifying skilled workers
- B. Acquiring skilled workers
- C. Training skilled workers
- D. Recruiting skilled workers

ANSWER: C

Descriptive analysis in HR analytic provides a view into activity, such as requisition volume, applicant or talent pool size, source of hires, etc.

- A. True

- B. False

ANSWER: A

Create a culture of transparency is one of the advantages of using HR analytics in talent acquisition.

- A. True

B. False

ANSWER: A

Which of the following is NOT a benefit of Business Analytics?

- A. Faster Decisions and Manage Risk**
- B. Better Customer Satisfaction and Reduce Cost**
- C. More Productive and Increase Profitability**
- D. High Network Speeds and Increase Data Storage**

ANSWER: D

Business Analytics projects require people from different disciplines and managerial levels.

A. True

B. False

ANSWER: A

Why is the understanding of data important?

- A. Make better decisions**
- B. Create new jobs**
- C. Become tech-savvy**
- D. Reduce work**

ANSWER: A

What kind of data does NOT help in understanding customer behaviours?

- A. Web browsing activity**
- B. Social media activity**
- C. Ecommerce buying activity**
- D. Credit card payment details**

ANSWER: D

Why are multidisciplinary teams required for data projects?

- A. Provide diverse perspective**
- B. Data analytics skills are hard to master**
- C. Not enough data professionals**
- D. Too expensive to only deploy data professionals**

ANSWER: A

Data by itself is not useful unless

- A. It is massive**
- B. It is processed to obtain information**
- C. It is collected from diverse sources**
- D. It is properly stated**

ANSWER: B

Organizations have hierarchical structures because

- A. it is convenient to do so**
- B. it is done by every organization**

- C. specific responsibilities can be assigned for each level
- D. it provides opportunities for promotions

ANSWER: C

Strategic information is needed for

- A. Day to day operations
- B. Meet government requirements
- C. Long term planning
- D. Short range planning

ANSWER: C

Which is the Application of Data Analytics

- A. Fraud Detection
- B. Risk Analysis
- C. Market research
- D. Data warehousing

ANSWER: A

Data Analysis is a process of?

- A. inspecting data
- B. cleaning data
- C. transforming data
- D. All of the given options

ANSWER : D

Which of the following is not a major data analysis approaches?

- A. Data Mining
- B. Predictive Intelligence
- C. Business Intelligence
- D. Text Analytics

ANSWER : B

The branch of statistics which deals with development of particular statistical methods is classified as

- A. industry statistics
- B. economic statistics
- C. applied statistics
- D. applied statistics

Answer : B

Which of the following is true about regression analysis?

- A. answering yes/no questions about the data
- B. estimating numerical characteristics of the data
- C. modeling relationships within the data
- D. describing associations within the data

Answer : C

Text Analytics, also referred to as Text Mining?

- A. True
- B. False
- C. Can be true or False
- D. Can not say

ANSWER : A

What is a hypothesis?

- A. A statement that the researcher wants to test through the data collected in a study.
- B. A research question the results will answer..
- C. A theory that underpins the study.
- D. A statistical method for calculating the extent to which the results could have happened by chance.

ANSWER : A

What is the cyclical process of collecting and analysing data during a single research study called?

- A. Interim Analysis
- B. Inter analysis
- C. inter item analysis
- D. constant analysis

ANSWER : A

The process of quantifying data is referred to as

- A. Topology
- B. Diagramming
- C. Enumeration
- D. coding

ANSWER : C

An advantage of using computer programs for qualitative data is that they _

- A. Can reduce time required to analyse data (i.e., after the data are transcribed)
- B. Help in storing and organising data
- C. Make many procedures available that are rarely done by hand due to time constraints
- D. All of the above

ANSWER : D

..... are the basic building blocks of qualitative data.

- A. Categories
- B. Units
- C. Individuals
- D. None of the above

ANSWER : A

This is the process of transforming qualitative research data from written interviews or field notes into typed text.

- A. Segmenting
- B. Coding
- C. Transcription
- D. Mnemoning

ANSWER : C

A graph that uses vertical bars to represent data is called a

- A. Line graph
- B. Bar graph
- C. Scatterplot
- D. Vertical graph

ANSWER :B

.....are used when you want to visually examine the relationship between two quantitative variables.

- A. Bar graph
- B. pie graph
- C. line graph
- D. Scatterplot

ANSWER : D

The denominator (bottom) of the z-score formula is

- A. The standard deviation
- B. The difference between a score and the mean
- C. The range
- D. The mean

ANSWER : A

Which of these distributions is used for a testing hypothesis?

- A. Normal Distribution
- B. Chi-Squared Distribution
- C. Gamma Distribution
- D. Poisson Distribution

ANSWER : B

A statement made about a population for testing purpose is called?

- A. Statistic
- B. Hypothesis
- C. Level of Significance
- D. Test-Statistic

ANSWER : B

If the assumed hypothesis is tested for rejection considering it to be true is called?

- A. Null Hypothesis

- B. Statistical Hypothesis
- C. Simple Hypothesis
- D. Composite Hypothesis

ANSWER : A

If the null hypothesis is false then which of the following is accepted?

- A. Null Hypothesis
- B. Positive Hypothesis
- C. Negative Hypothesis
- D. Alternative Hypothesis.

ANSWER : D

Alternative Hypothesis is also called as?

- A. Composite hypothesis
- B. Research Hypothesis
- C. Simple Hypothesis
- D. Null Hypothesis

ANSWER : B

Data by itself is not useful unless

- A. It is massive
- B. It is processed to obtain information
- C. It is collected from diverse sources
- D. It is properly stated

ANSWER: B

Decision support systems are used for

- A. Management decision making
- B. Providing tactical information to management
- C. Providing strategic information to management
- D. Better operation of an organization

ANSWER: A

Business analytics (BA) is a broad category of application programs which includes :

- A. Decision support
- B. Data mining
- C. OLAP
- D. OLAB

ANSWER: B

Decision support systems are used by

- A. Line managers.
- B. Top-level managers.
- C. Middle level managers.
- D. System users

ANSWER: B

Which of following is not phase of decision making process

- A. Design
- B. Analysis
- C. Intelligence
- D. Choice

ANSWER: B

.....is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions

- A. best practice
- B. data mart
- C. business information warehouse
- D. business intelligence

ANSWER: D

Organizations have hierarchical structures because

- A. it is convenient to do so
- B. it is done by every organization
- C. specific responsibilities can be assigned for each level
- D. it provides opportunities for promotions

ANSWER: C

Strategic information is needed for

- A. Day to day operations
- B. Meet government requirements
- C. Long range planning
- D. Short range planning

ANSWER: C

Decision support systems are essential for

- A. Day-to-day operation of an organization.
- B. Providing statutory information.
- C. Top level strategic decision making.
- D. Ensuring that organizations are profitable.

ANSWER: C

Computer information system are most successful in providing information for:

- A. Control decisions
- B. Planning decision
- C. Strategic decision
- D. None of the above

ANSWER : A

Which is the last Phases of mathematical models for decision making

- A. Problem Identification**
- B. Implementation and Testing**
- C. Model Formation**
- D. Development of Algorithm**

ANSWER : C

Which mathematical model aims at understand the mechanisms that regulate the development of intelligence, ability to extract knowledge from past experience in order to apply it in the future.

- A. Risk analysis models**
- B. Optimization models**
- C. Pattern recognition Models**
- D. Waiting line models**

ANSWER: A

In which Mathematical mode the decision maker is required to choose among a number of available alternatives.

- A. Risk analysis models**
- B. Optimization models**
- C. Pattern recognition Models**
- D. Waiting line models**

ANSWER:B

Which is not the Phase of data Mining Process

- A. Data Gathering**
- B. Selection of Attributes**
- C. Prediction and interpretation**
- D. Data Discarding**

ANSWER: D

Data Inception Means

- A. inspection of each missing value**
- B. identify missing values**
- C. replacement of missing Data**
- D. discard all records**

ANSWER: A

Data imputation Means

- A. inspection of each missing value**
- B. identify missing values**
- C. replacement of missing Data**
- D. discard all records**

ANSWER: C

Data mining activities can be subdivided into two major investigation streams , which are

- A. Interpretation and Sampling**
- B. Interpretation and Prediction.**
- C. Forecast and Prediction**
- D. Forecast and Interpretation**

ANSWER: A

Which is the Application of Data Analytics

- A. Fraud Detection**
- B. Risk Analysis**
- C. Market research**
- D. Data warehousing**

ANSWER: A

.....learning analyses are not guided by a target attribute.

- A. Supervised**
- B. Guided**
- C. Unguided**
- D. Unsupervised**

ANSWER: A

Point out the correct statement?

- A. In R, a function is an object which has the mode function**
- B. R interpreter is able to pass control to the function, along with arguments that may be necessary for the function to accomplish the actions that are desired**
- C. Functions are also often written when code must be shared with others or the public**
- D. All of the mentioned**

Answer: D

What is true about Data Visualization?

- A. Data Visualization is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts.**
- B. Data Visualization helps users in analyzing a large amount of data in a simpler way.**
- C. Data Visualization makes complex data more accessible, understandable, and usable.**
- D. All of the above**

Answer : D

Data can be visualized using?

- A. graphs**
- B. charts**
- C. maps**
- D. All of the above**

ANSWER : D

Data visualization is also an element of the broader

- A. deliver presentation architecture**

- B. data presentation architecture
- C. dataset presentation architecture
- D. data process architecture

Answer : B

Which method shows hierarchical data in a nested format?

- A. Treemaps
- B. Scatter plots
- C. Population pyramids
- D. Area charts

ANSWER : A

Which of the following is false?

- A. data visualization include the ability to absorb information quickly
- B. Data visualization is another form of visual art
- C. Data visualization decrease the insights and take solver decisions
- D. None Of the above

ANSWER : C

Common use cases for data visualization include?

- A. Politics
- B. Sales and marketing
- C. Healthcare
- D. All of the above

ANSWER : D

Which of the following plots are often used for checking randomness in time series?

- A. Autocausation
- B. Autorank
- C. Autocorrelation
- D. None of the above

ANSWER : C

Data Analysis is a process of?

- A. inspecting data
- B. cleaning data
- C. transforming data
- D. All of the above

ANSWER : D

Which of the following is not a major data analysis approaches?

- A. Data Mining
- B. Predictive Intelligence
- C. Business Intelligence
- D. Text Analytics

ANSWER : B

How many main statistical methodologies are used in data analysis?

- A. 2
- B. 3
- C. 4
- D. 5

ANSWER : A

In descriptive statistics, data from the entire population or a sample is summarized with ?

- A. integer descriptors
- B. floating descriptors
- C. numerical descriptors
- D. decimal descriptors

ANSWER : C

Data Analysis is defined by the statistician?

- A. William S.
- B. Hans Peter Luhn
- C. Gregory Piatetsky-Shapiro
- D. John Tukey

ANSWER : D

Which of the following is true about hypothesis testing?

- A. answering yes/no questions about the data
- B. estimating numerical characteristics of the data
- C. describing associations within the data
- D. modeling relationships within the data

ANSWER : A

The goal of business intelligence is to allow easy interpretation of large volumes of data to identify new opportunities.

- A. TRUE
- B. FALSE
- C. Can be true or false
- D. Can not say

ANSWER : A

The branch of statistics which deals with development of particular statistical methods is classified as

- A. industry statistics
- B. economic statistics
- C. applied statistics
- D. applied statistics

ANSWER : D

Which of the following is true about regression analysis?

- A. answering yes/no questions about the data
- B. estimating numerical characteristics of the data
- C. modeling relationships within the data
- D. describing associations within the data

ANSWER : C

Text Analytics, also referred to as Text Mining?

- A. TRUE
- B. FALSE
- C. Can be true or false
- D. Can not say

ANSWER : A

Data Analysis is a process of?

- A. inspecting data
- B. cleaning data
- C. transforming data
- D. All of the above

ANSWER : D

Which of the following is not a major data analysis approaches?

- A. Data Mining
- B. Predictive Intelligence
- C. Business Intelligence
- D. Text Analytics

ANSWER : B

How many main statistical methodologies are used in data analysis?

- A. 2
- B. 3
- C. 4
- D. 5

ANSWER : A

In descriptive statistics, data from the entire population or a sample is summarized with ?

- A. integer descriptors
- B. floating descriptors
- C. numerical descriptors
- D. decimal descriptors

ANSWER : C

Data Analysis is defined by the statistician?

- A. William S.
- B. Hans Peter Luhn
- C. Gregory Piatetsky-Shapiro

D. John Tukey

ANSWER : D

Which of the following is true about hypothesis testing?

- A. answering yes/no questions about the data
- B. estimating numerical characteristics of the data
- C. describing associations within the data
- D. modeling relationships within the data

ANSWER : A

The goal of business intelligence is to allow easy interpretation of large volumes of data to identify new opportunities.

- A. TRUE
- B. FALSE
- C. Can be true or false
- D. Can not say

ANSWER : A

The branch of statistics which deals with development of particular statistical methods is classified as

- A. industry statistics
- B. economic statistics
- C. applied statistics
- D. applied statistics

ANSWER : D

Which of the following is true about regression analysis?

- A. answering yes/no questions about the data
- B. estimating numerical characteristics of the data
- C. modeling relationships within the data
- D. describing associations within the data

ANSWER : C

Text Analytics, also referred to as Text Mining?

- A. TRUE
- B. FALSE
- C. Can be true or false
- D. Can not say

ANSWER : A

What is true about Machine Learning?

- A. Machine Learning (ML) is that field of computer science
- B. ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method.
- C. The main focus of ML is to allow computer systems learn from experience without being explicitly programmed or human intervention.

D. All of the above

ANSWER : D

ML is a field of AI consisting of learning algorithms that?

- A. Improve their performance**
- B. At executing some task**
- C. Over time with experience**
- D. All of the above**

ANSWER : D

A begins by hypothesizing a sentence (the symbol S) and successively predicting lower level constituents until individual preterminal symbols are written.

- A. bottom-up parser**
- B. top parser**
- C. top-down parser**
- D. bottom parser**

ANSWER : C

A model of language consists of the categories which does not include

- A. System Unit**
- B. structural units.**
- C. data units**
- D. empirical units**

ANSWER : B

Different learning methods does not include?

- A. Introduction**
- B. Analogy**
- C. Deduction**
- D. Memorization**

ANSWER : A

The model will be trained with data in one single batch is known as ?

- A. Batch learning**
- B. Offline learning**
- C. Both A and B**
- D. None of the above**

Answer : C

Which of the following are ML methods?

- A. based on human supervision**
- B. supervised Learning**
- C. semi-reinforcement Learning**
- D. All of the above**

ANSWER : A

In Model based learning methods, an iterative process takes place on the ML models that are built based on various model parameters, called?

- A. mini-batches
- B. optimizedparameters
- C. hyperparameters
- D. superparameters

ANSWER : C

For Internal Circulation at BVIMR only

For Internal Circulation & BVIMR only

PREVIOUS YEAR SAMPLE QUESTION PAPER

Subject : Business Intelligence

Day : Wednesday
Date : 12/04/2017



Time : 10.00 AM TO 01.00 PM
Max Marks : 100 Total Pages : 1

N.B.:

- 1) Attempt **ANY FOUR** questions from Section – I and **ANY TWO** questions from Section – II.
- 2) Answers to both the sections should be written in **SEPARATE** answer books.
- 3) Figures to the right indicate **FULL** marks.

SECTION – I

- Q.1 What is decision support system? Explain its role in decision making process. [15]
- Q.2 Explain the data warehouse architecture in detail with all its components. [15]
- Q.3 Why metadata is important? Explain with suitable example. [15]
- Q.4 Differentiate between OLAP and OLTP giving suitable examples for each point. [15]
- Q.5 What is the need of data mining? Explain association rule mining in detail. [15]
- Q.6 Data mart is essential? Explain in brief. What are the different approaches to build the data marts? [15]
- Q.7 What is the need of clustering? Explain major clustering method. [15]

SECTION – II

- Q.8 What is the difference between classification and prediction? Compare different classification methods. [20]
- Q.9 What is data pre-processing? Explain all the data pre-processing techniques. [20]
- Q.10 How KDD works, explain in detail by taking suitable example? [20]

* * * *

Notes

Topic:.....

Date:.....

Notes

Topic:.....

Date:.....

Notes

Topic:.....

Date:.....

Notes

Topic:.....

Date:.....