

BUSINESS INTELLIGENCE & DATA MINING: TECHNIQUES & APPLICATIONS

Independent Study (CS 597-10)
Instructor – Prof. Leslie Lander



BINGHAMTON | THOMAS J. WATSON SCHOOL OF
UNIVERSITY | ENGINEERING AND APPLIED SCIENCE

Submitted by-
Bhuvan Chadha
(B No. - B00815783)

Introduction

Analytics

The term 'Analytics' is broadly employed in businesses for data-driven analyses used within the decision-making method. The term refers to the appliance of various strategies of study of knowledge to resolve issues. a group of 'data science' processes and techniques square measure used along to get data, analyze it, forecast numerous attainable actions towards an answer, and quantify the impact of those possible selections of the longer term on optimizing business.

With knowledge Analytics (analytics-driven knowledge analyses), enterprises square measure able to get a '360 degree' read of their operations and customers. The insight so gained is helpful in directional and optimizing, and automating their deciding to fulfil their structure goals. Business Analytics (data-driven analyses of business data) helps in strategic, operational and plan of action deciding through the business verticals together with medium, retail and money services.

Business Intelligence (Insight into the past)

A group of architectures and technologies capable of reworking data into data that may be used for business functions by creating it meaning, is named Business Intelligence (BI). exploitation basic metallic element tools tools, businesses will simply interpret immense volumes of information, and live past performance.

The term Business Intelligence is usually comprehensive of the creation and maintenance of information Warehouses. It's essential to gather and integrate the varied/diverse knowledge kept in varied primary and secondary sources, most of that area unit a part of operational systems. knowledge warehouses gather and merge along knowledge from transaction-processing systems, with its own information.



Figure 1. Phases of Business Intelligence

Predictive Analytics (Understanding the future)

Predictive Analytics goes beyond BI by using subtle quantitative techniques (for example, descriptive modeling, prophetic modeling, simulation, and optimization) to supply insights that ancient approaches to metal (such as question, reporting, outline statistics, information visualization) square measure unlikely to get. It deals with automatic discovery of meaning patterns in structured further as unstructured information. prophetic Analysis is capable of predicting what will happen. These analytics relating to the understanding of the long run, and provide empirical insights to firms.

The main subfields of prophetic analysis square measure descriptive modeling, prophetic modeling, and improvement and simulation. Even on undefeated adoption of those analytical techniques, the choice is taken by decision-makers, on the idea of informal and unstructured info offered to change and adopt the recommendations and conclusions mistreatment analytics.

Machine Learning, Data Mining and Predictive Analysis

Methodologies and technologies from each statistics and computing have vie a crucial role within the development of prophetic analysis. the most contributions to the present discipline come back from Machine Learning and data processing. The set of techniques for the extraction of prophetic analysis models from information is currently called Machine Learning.

Data Mining springs from Machine Learning as a hunt field, that focuses on considerations raised in analysis on real-world applications. As such, analysis targeting business applications and business-related issues of knowledge analysis tend to drift within the direction of knowledge Mining. each the fields area unit associated with one another.

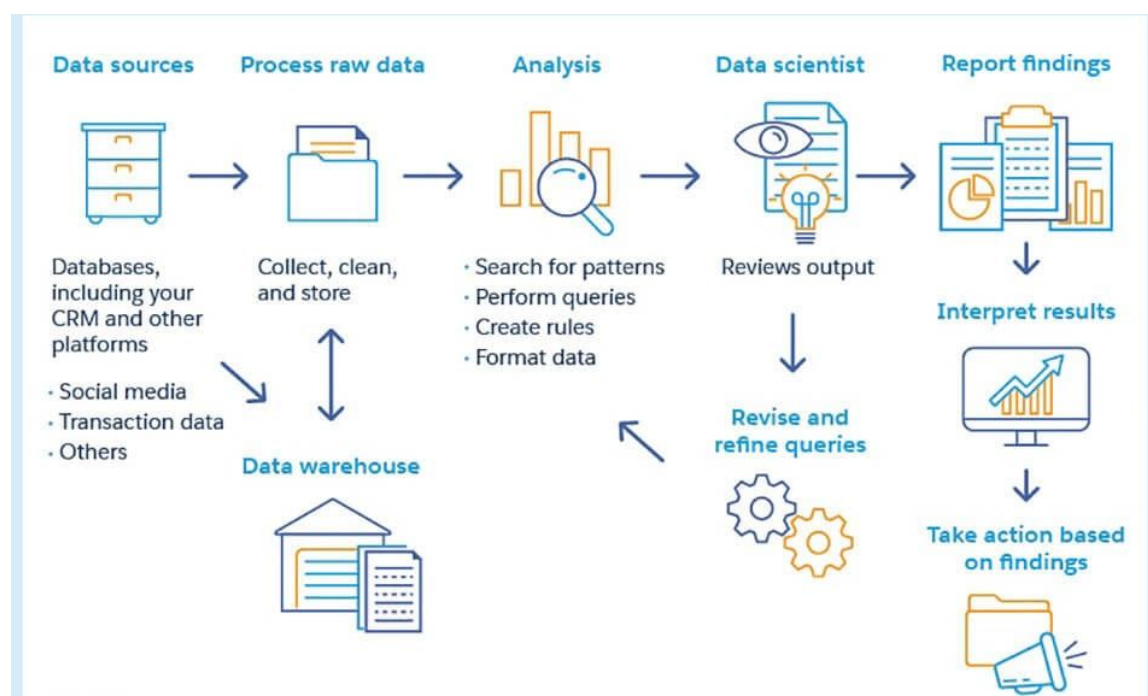


Figure 2. A flowchart showing how data Mining Works

Basic Data Analytical Techniques

The following are the commonly used basic techniques in analysis -

- Data Warehousing
- Naive Bayes Classifier, k-Nearest Neighbour algorithm, linear and logistic regression
- Support vector Machines
- Neural Networks
- Fuzzy Inference Systems
- Decision Trees
- Data Clustering
- Data Transformations
- Optimization Algorithms
- Learning Associations

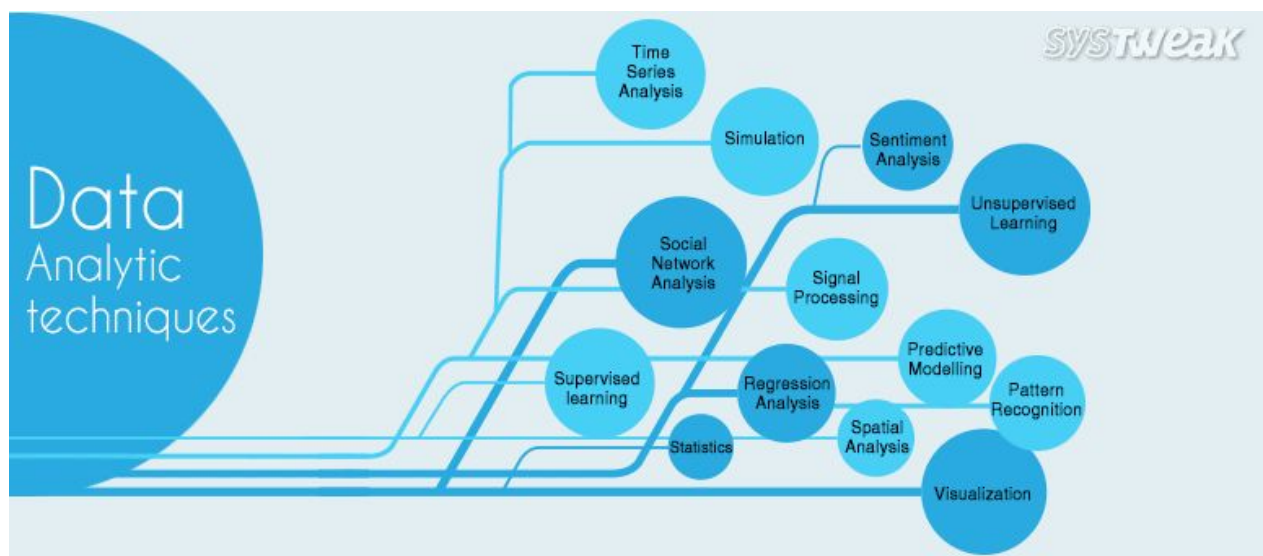


Figure 3. Big Data Analytics Techniques

In addition to trade verticals, money Services Analytics, Sales Analytics, client Relationship Management (CRM) analytics, medium analytics, producing analytics, etc, recent advances have semiconductor diode to the most recent and hottest trends in business analytics - *Text Analytics* and *multimedia system Analytics*.

- **Text Analytics** - permits users to derive insights from unstructured text knowledge collections. Unstructured text knowledge is remodeled into a structured format that may be used because the base for prophetic and descriptive analysis.
- **Multimedia Analysis** - Uses varied technologies to derive insights from unstructured knowledge that comes along in multiple modalities as well as pictures, audio, video, and text.

The focus of discussions touching on numerous aspects of a company's product portfolio these days is shifting from individual company websites to blogs, forums, and alternative cooperative sites - in alternative words, the social media. This new media permits everybody to post comments and opinions regarding corporations and their merchandise, which can impact the perception and shopping for behavior of numerous consumers. promoting

corporations are involved regarding this as a result of it's difficult to envision the unfold of negative info, neither is it simple to search out or establish identical in giant areas, like forums, blogs and social networking sites.

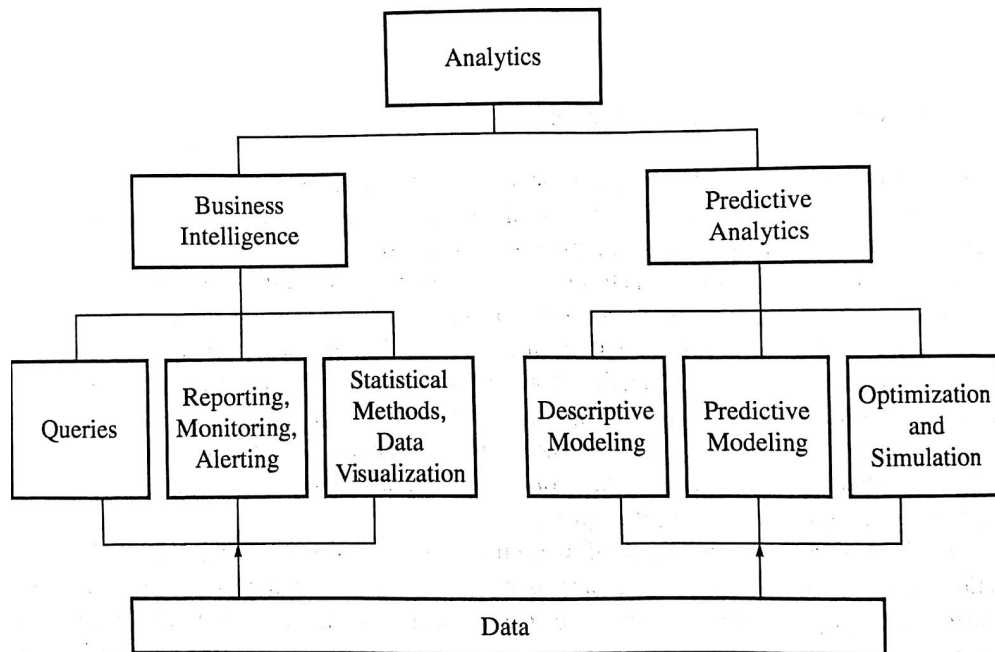


Figure 4. Analytics and relevant sub-fields

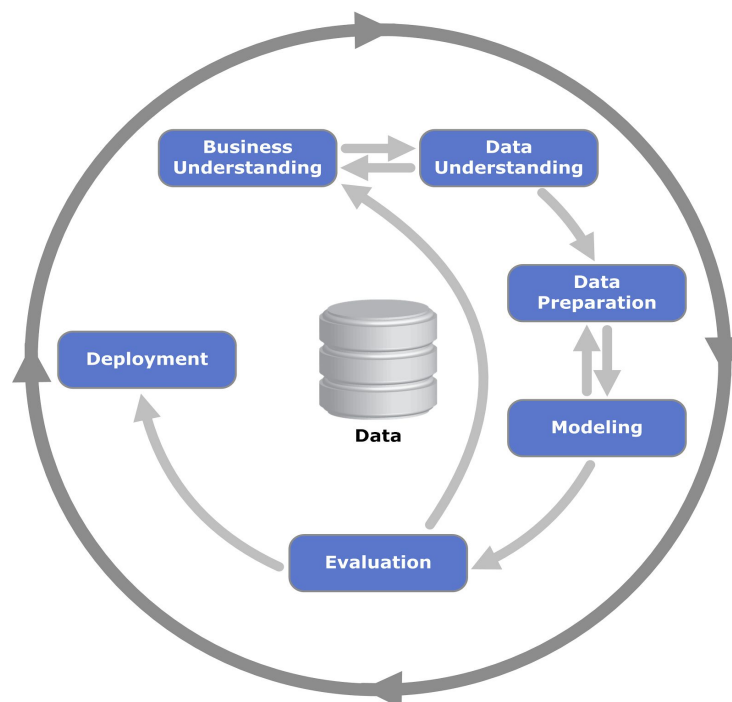


Figure 5. The Cross Industry Standard Process For Data Mining (CRISP-DM) Model

Data Warehousing: General Architecture & OLAP Operations

Data Warehouses often adopt a three-tier architecture -

The Bottom Tier

It is a data warehouse server that's typically a relational database accessed through some variant of SQL. The central repository is the heart of the information warehouse.

The origin of information lies within the supply systems, that area unit operational systems, and external information feeds. Their style ensures potency operational, not call support. information warehouse systems extract information, clean information and additionally perform information integration and transformation victimisation back-end devices/utilities. Back-end devices and utilities perform load and refresh functions additionally so as to shift information from supply systems into the info warehouse or the lowest level.

The Middle Tier

The Middle Tier is an OLAP server executed with the utilization of a multidimensional information model. This methodology views the data as a *data cube*. Cubes are conveniently stored in relational databases with the help of a data structure known as the *star schema*.

A data cube permits multidimensional views of the data and is defined by *dimensions* and *facts*. Generally speaking, dimensions are entities or viewpoints according to which an organization wishes to maintain records. These dimensions facilitate the tracking of periodic sales of products, branch wise or location wise.

The Top Layer

It is a front-end client layer which contains query and reporting tools, analysis tools, etc.

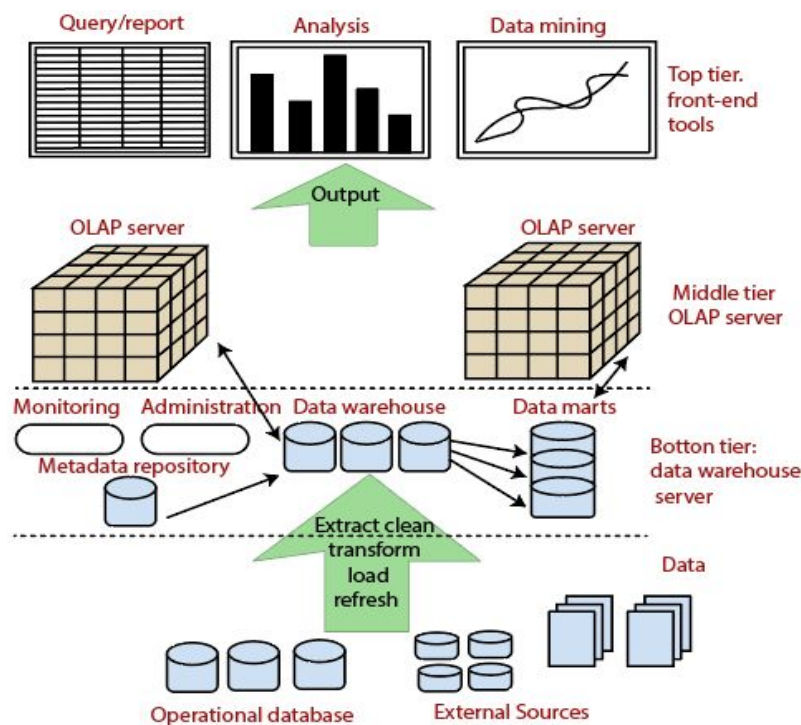


Figure 6. Three-Tier Data Warehouse Architecture

Technologies For Big Data

Big Data is often referred to as the data that exceeds the capability of widely-used hardware environments and software tools that capture, manage, process, it within an acceptable time period for its users. As per the definition, the qualification of big data will alter with technological advancement. Big data definitely involves huge volumes of data. It also possesses high *velocity* (rate of transmission and receipt of data), and a *variety* of data sources compared to conventional data sources. While the focus is mostly on its size, the challenging part of big data actually is the absence of structure in it as compared to conventional data, which is structured and possesses a fixed file format. *Web Data* can be said to be the most commonly used and recognized source of big data. There are a number of sources of enormous information, with their own significant users. These are-

- Web Data
- Text Data
- Social Network Data
- Time and Location DataSensor data

Clearly, Big Data could be all new sources of data and not only an extended collection of old and traditional data. It is, at times, generated mechanically employing a machine in an automatic manner. A device embedded in Associate in Nursing engine, for eg, throws out information to its surroundings although no-one demands for it. We tend to seize all this information and solely hassle regarding it throughout the analysis stage.

Analytics professionals have used a variety of tools & technologies over the years that have created it doable to arrange information for analysis, execution of analytic algorithms and assessment of the results.

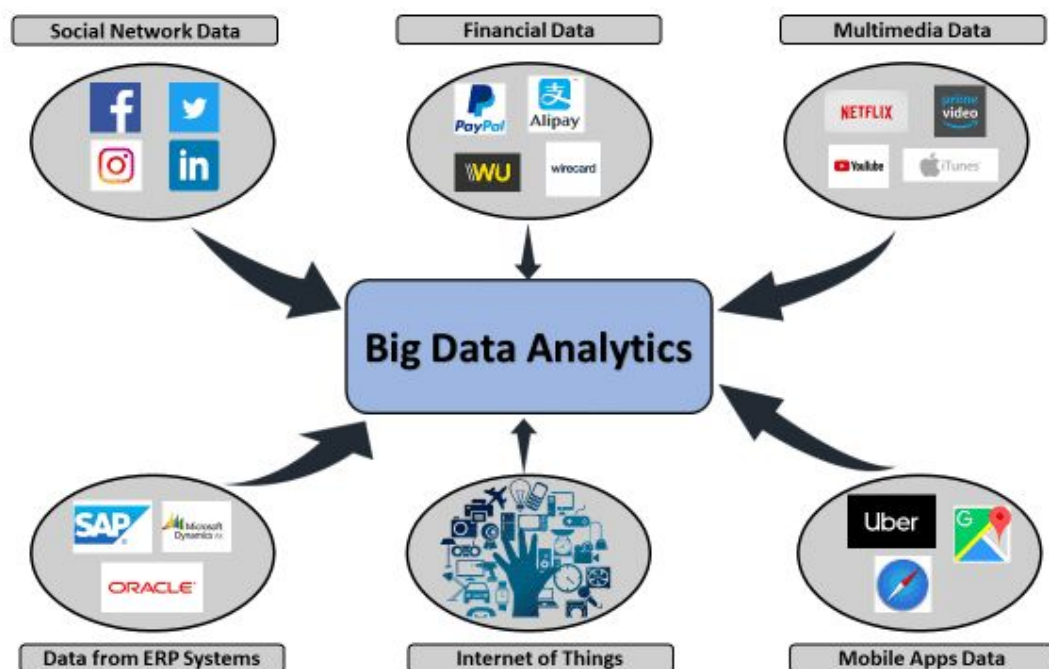


Figure 7. Sources of Big Data

Data Mining Applications and Trends

Some of the Data Mining Applications are as follows-

- Financial Services
- The Retail Industry
- The Telecommunication Industry
- The Insurance Industry
- Healthcare Services
- Supply Chain Management
- e-Governance
- Customer Relationship Management (CRM)

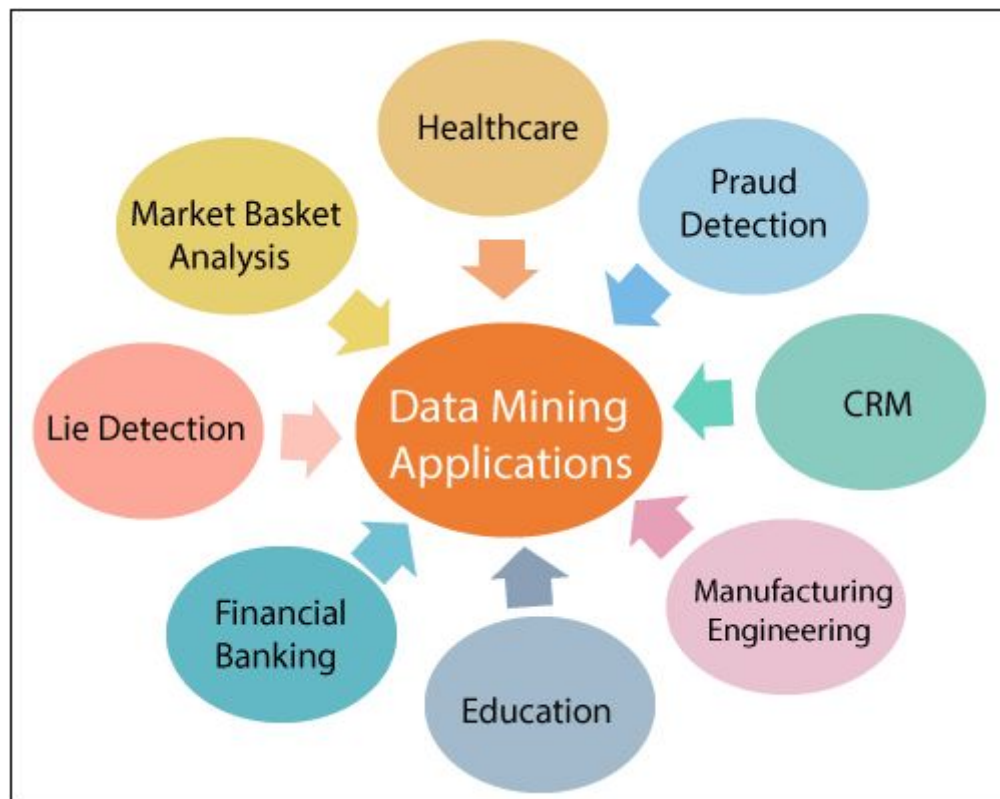


Figure 8. Data Mining Applications

Some of the trends in data mining applications are as follows that reflect the pursuit of the real time challenges -

- Mining Sequence Data
- Mining Graphs and Networks
- Mining Spatio Temporal Data
- Mining Cyber-Physical Data
- Mining Multimedia Data
- Mining Web Data
- Mining Data Streams
- Recommender Systems
- Adversarial Situations
- Data Mining for Big Data

Conclusion

In conclusion, knowledgeable decisions to grow an organization's overall performance is very important. Business Intelligence and Big Data Technology have all the required tools that benefit a company. Although, earlier, it was viewed for years that BI tools are only accessible to larger organizations, the belief has changed regularly and almost all the industries, however big or small, are establishing to make use of the magical tool - Business Intelligence.

References

- <https://www.mastersindatascience.org/resources/what-is-data-analytics>
- <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>
- <https://www.oracle.com/a/ocom/docs/machine-learning-goes-to-the-cloud-ebook.pdf>
- https://en.wikipedia.org/wiki/Big_data
- Applied Machine Learning, M. Gopal