

Q3. Data mining techniques →

Businesses that utilize data mining are said to have competitive advantage, ~~over~~ better understanding of their customers, good oversight of business operations, new business opportunities etc.

Some businesses are looking for best ways to get new customers, some are looking for new marketing techniques, some are working to improve their systems.

The data mining process is what ~~tells~~ gives businesses the understanding of how to make their decisions, analyse their information and move forward.

★ Data mining techniques in business analytics →

1. Classification

- This data mining technique is more complex
- Involves using attributes of data to move them into discernable categories

1. Clustering -

- Clustering is a technique used to represent data visually such as graphs representing buying trends or sales demographics for a particular product.
- Clustering refers to the process of grouping a series of different data points based on their characteristics. By doing so, miners can seamlessly divide data into subsets.

→ Methods for data clustering -

- Partitioning method
- Hierarchical method
- Density based method
- Grid based method
- Model based method

2. Association -

- Association rules are used to find associations, correlations etc. between points in a data set.
- Data miners use association to discover unique or interesting relationships between variables in databases.
- Two approaches -
 - Single-dimensional association - involves looking for one repeating instance of a data point or attribute.
 - Multi-dimensional association - involves looking for more than one data point in a data set.

3. Data Cleaning -

- It is the process that prepares the data to be mined.
- Data cleaning involves organizing data, elimination of duplicates or corrupted data, filling in any null values etc.
- When this process is completed, the most useful information can be harvested for analysis.

4. Data Visualization -

- Data visualization is the translation of data into graphic form to illustrate its meaning to business stakeholders.
- Data can be presented in visual ways through graphs, charts, maps, diagrams and more.
- This is the primary way in which data scientists display their findings.

5. Classification -

- It is the process by which data points from large data sets are assigned to categories based on how they are used.
- In data mining, classification is considered to be a form of clustering. Classification is used to designate broad groups within a demographic, target audience or user base through which businesses can gain stronger insights.
- Methods include - Logistic regression, decision trees, K-nearest neighbour (knn), Naive Bayes, Support Vector Machine (SVM) etc.

6. Machine Learning -

- Machine learning is the process by which ~~the~~ computers use algorithms to learn on their own.

- Machine learning applications in data mining are vast and they both fall under the umbrella of data science.
- Methods include
 - Supervised learning
 - unsupervised learning
 - reinforcement learning
 - semi supervised learning

7. Neural networks →

- Artificial neural networks attempt to mimic the way human brain operates.
- Neural networks combine multiple computer processors (similar to the way brain uses neurons) to process data, make decisions and learn the way a human would, ^{or} at least as closely as possible.
- Neural networks can be used to predict consumer buying patterns and focus marketing campaigns on specific demographics.

8. Outlier detection →

- Outlier detection is a key component of maintaining safe databases.
- Outlier detection looks for unique data points that differ from the rest or diverge from the overall sample.
- Companies use it to test for fraudulent transaction, such as abnormal credit card usage, which may indicate theft.

9. Prediction →

- Predictive modelling seeks to turn data into ^aprojection of future action or behaviour.
- These models examine data to find trends or patterns, and then calculate the probabilities of a future outcome.
- Methods include →
 - Forecast modelling
 - Cluster modelling
 - Classification modelling
 - Time series modelling

Unit - 3

Q1. Data mining and Text mining

Data mining process →

1. Defining the problem -

Data mining starts with clearly defining the business problem. eg. increase sales, get more return customers etc.

2. Selecting features / variables -

Business can collect data based on the customers and what they purchased from the company to examine returning customers and create customer profiles, eg. customer age, location, income would be helpful variables.

3. Collecting and curating data -

Once the question and dataset are determined, the data engineers can create the data pipeline to ^{collect} obtain the data and put the existing data into required format. Dataset is curated to give insight about this particular business problem.

4. Analyzing the data -

Data scientists investigate the data to remove outliers or anomalies and analyze it to determine patterns to solve business problems.

5. Make business decisions -

Once the results are out, BA team can make data driven decisions to change or optimize a certain business strategy or operation.

6. Track changes -

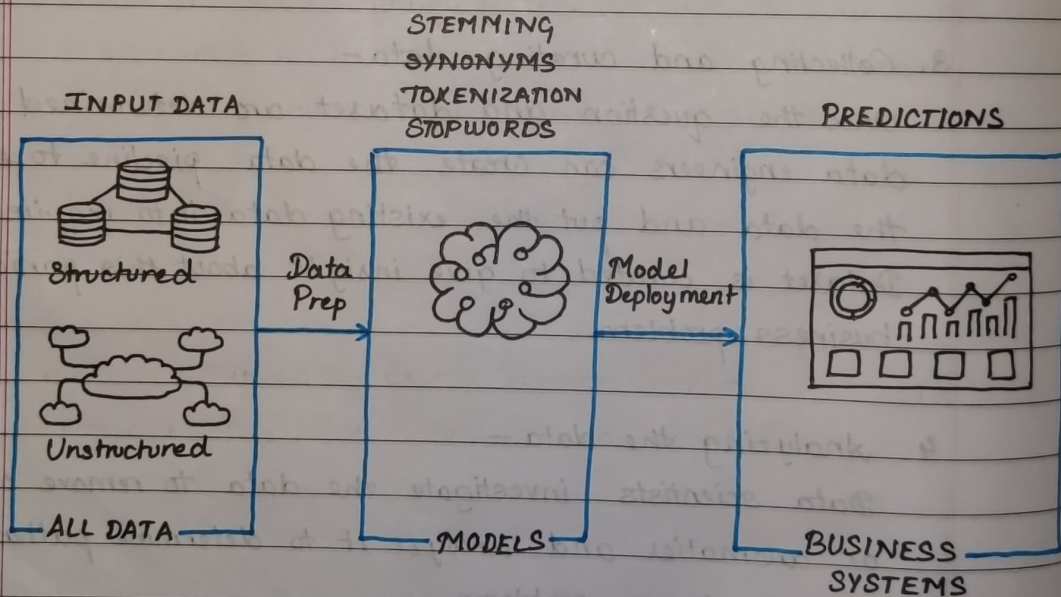
data collection and analysis process continues to see if the ~~changes~~ decisions work as expected.

7. Adjust and Repeat.

Text analytics and Text mining →

Researchers in this stream specialize in the use of data mining and statistical machine learning to analyse structured and unstructured data.

Text analytics is focussed on extracting information from unstructured text to form structured data patterns.



Text mining and text analytics overview

Benefits of text analytics →

- It helps businesses understand customer trends, product performance etc. This results in quick decision making, increased productivity, cost savings etc.

- It assists in understanding general trends and opinions in society that enable governments and political bodies to make decisions.
- It helps search engines and information retrieval services to improve their performance and provide fast user experience.

Sentiment analysis →

It is used to identify the emotions conveyed by unstructured text. The input text involves product reviews, customer interactions, social media posts, blogs etc.

Polarity analysis identifies if the text expresses +ve or -ve sentiment, while the categorization technique involves more fine grained analysis of emotions (confused, dissatisfied, angry etc.). Sentiment analysis is used to measure customer response to a product/service and track how customer sentiment evolves over time.

Text mining applications →

Sentiment analysis is a widely used text mining application that can track customer sentiment about a company.

Such information can be used to fix product issues, improve customer service etc.

Other common text mining uses include screening job candidates based on the wording in their resumes, blocking spam mails, classifying website content, flagging insurance claims as fraudulent, analyzing descriptions of medical symptoms to aid in diagnoses etc.

Benefits of text mining →

- Using text mining can help companies detect product and business problems and address them before they become big issues that affect sales.
- Mining text in customer reviews can help identify desired new features which can strengthen product offerings.
- Fraud detection, risk management, web content management etc. are other functions that can benefit from the use of text mining tools.
- In healthcare, technology can be able to help diagnose illness and medical conditions in patients based on the symptoms they report.

Text mining challenges and issues →

Text mining can be challenging because the data is often inconsistent, vague and contradictory.

Efforts to analyse it are further complicated by ambiguities that result from differences in syntax and semantics, regional dialects, slang, sarcasm etc.

The deep learning models used for text mining require large amounts of training data which can make them expensive to run.