Name - Amod Dhopavkar
Class - TE11
Roll No - 33304

DATE __/__/____
PAGE NO _____

mycompanion

DSBDA UT-2

Q.1) a) Explain in detail different steps in data analytics project life cycle? List different libraries used in R.

Ans) 1. Discovery : Stakeholders constantly analyze the business trends similar case studies and domain of the business industry.

2. Data preparation : Once the discovery is complete with a walkthrough of the business model, metrics and results the data aggregation occurs during preparation phase.

3. Model planning : The data science team includes the stakeholders and build the required model

4. Communication result : Another vital methodology involving all the stakeholders to build a data driven organisation by infusing data analytics culture.

5. Operationalize : Stakeholders operationalize results from sandbox and plan to deploy the data science project

Q.1) b) What kinds of issues affect the quality of data?

Ans) 1. Entry quality : Did the information enter the system correctly at source.

2. Process quality : Proper quality control at each touchpoint along the path can help ensure the problems are rooted out.

3. Interaction quality : All known information about an object integrated to provide accuracy.

4. Usage quality : Is the information used and interpreted correctly at the point of access.

Q.2) a) Discuss different types of data visualisation techniques.

Ans) The ever growing volume of data and its importance to businesses make data visualisation an essential part of companies.

1. Chart : Easiest way to show development of one or more datasets

2. Plots : Allow to distribute 2 or more datasets over a 2D or a 3D space

3. Maps : The allow to position elements on relevant objects

4. Diagrams and notices : Used to demonstrate complex data relationship

Q.2) b) What are the challenges in visualising big data?

Ans) 1. The over simplification of data is one of the biggest drawbacks of visualisation

2. The human limitation of algorithm. This is the biggest potential problem and also the most complicated

3. Over-reliance on visuals :
This is more of a problem with the consumers than the developers, but it underlines the potential impact of visualisation in general.

4. The inevitability of visualisation

There are many tools available to help us understand complex data set with visual diagram, chart and illustration

33304

Q.3)a) What is text mining? Draw and explain text mining architecture.

Ans) Text mining is also known as text data mining. The purpose is to extract meaningfull information from unstructured text.

Process of text mining →

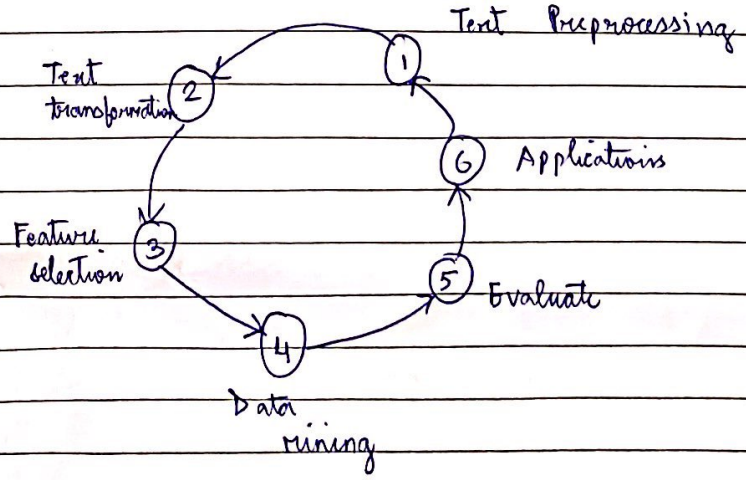
1. Preprocessing : It involves a series of steps as given below
   i) Text cleanup
   ii) Tokenization
   iii) Part of speech tagging

2. Text Transformation : The text document is represented by the words it contains and their occurrences

3. Feature selection : It is also known as variable selection

4. Data mining : The text mining process merges with the traditional process

5. Evaluate : Evaluate the result.



Q.3)b) How mobile analytics is different than social media analytics ? Explain with example.

Ans) The term mobile analytics is defined as it is used for capturing the data coming from mobile app, web app and websites to assist companies get better engagement, retention and conversion.

Mobile analytics provides companies with invaluable insights into the usage of the end consumers. It is useful for tracking distinct users, record their demographics and behaviour.

The tracking technology differs amongst websites which use a Software Development Kit (SDK). The SDK uses Javascript or cookies. ~~The data can be~~ The analytics platform is run every time a visitor takes any action.

Tools for Mobile Analytics →

1. Google Analytics
2. AI Internet
3. Adobe Analytics
4. Smartlook
5. Mixpanel
6. Webtrends Analytics