**Engineering Data Mining Discussion Questions**

Data mining is the process by which patterns, correlations, and anomalies are retrieved from data sets to come up with important patterns (Zhang et al., 2019). Decision-makers can use the information retrieved from data mining to increase production within their organizations, reduce costs and improve customer relationships. Text mining is a form of data mining that is specifically applied to unstructured text files (Zhang et al., 2019). Text mining involves analyzing a large amount of data by structuring text inputs and sentimental analysis to find meaningful patterns and information for the decision-making process. The sentimental analysis utilizes opinion or source to identify and classify data making it a form of data mining that primarily deals with data labeling (Zhang et al., 2019).

Text mining is the use of data mining technics to manage unstructured and semi-structured text files (Parvathi, 2021). The difference is that text mining analyzes words, whereas data mining primarily deals with numerical data. Text mining is gaining meaning in the current technological sphere due to the presence of a large volume of text data being generated on a day-to-day operation (Parvathi, 2021). Application of text mining includes text analysis in textual industries like the law courts, research centers, technology hubs, marketing, and finance sectors (Parvathi, 2021). The applications include;

* Extracting information by identifying keywords and relationships following designed sequences and patterns.
* Tracking topics based on user profile, documents, and interests.
* Clustering and categorizing data based on predefined patterns, topics, and themes.
* Matching patterns to find the answer to different questions.
* Connecting related documents and texts based on the topics and themes to help find information easily the users.

Natural language processing is part of artificial intelligence that is connecting the two topics by helping in text mining (Deng & Liu, 2018). It studies and understands human natural language to help in processing text data into more informed patterns and formal representations using models and numerical (Deng, & Liu, 2018). NLP uses computer programs to help in manipulating textual data to retrieve patterns and sequences. Text mining applies natural language to induce structure in collecting data and categorizing data into clusters, classes, sequences, and associations to retrieve patterns and knowledge using the data mining algorithm (Deng, & Liu, 2018).

The main limitations of using natural language processing in text mining include an error in text or speech, contextual words and phrases, lack of research and development, and ambiguity of some words.

**Exercise 3**

Due to advancing technology and increasing internet speed, a lot of data is generated following daily activities (Chen, & Tu, 2019). This has made platforms like eBay experience real data challenges. However, data is essential since it helps the platform to evaluate services offered to their clients.

eBay is using analytics to study and evaluate its customers to understand them better. The data analysis is helping the platform to personalize engagements with its customers. It also helps understand customer preferences, learn existing clients and engage clients at a personal level using their preference.

Data analysis is a challenge to even the eBay platform (Chen, & Tu, 2019). Their web analytics program is like a recording device planted in every client, and it is able to generate tones of data that is challenging to handle (Chen, & Tu, 2019). For example, the eBay platform collects data generate for around 100 million hours in a month. Digging out patterns to help understand customers and get more data using data mining tools is essential.

**Internet exercise 7**

Classification software. It is a multiple approach software that utilizes decision tree and neural network models o combine and compare data.

Clustering and segmentation software. Uses Bayesian networks and algorithms to cluster and segment data, respectively. It consists of applications like BayesiaLab, ClustanGraphics3, CMSR Data Miner IBM SPSS modeler, CLUTO, ELKI, plus many other features.

Web Usage Mining Software. It consists of processes like A1WebStats, 11Ants Model Builder, ClickTracks, Download Analyzer, plus many other processes found within the software.