# **Activity 1**

## **Advance Data Analytics, NAS2001, E21**

**NAME ⇒ Abhishek Srivastava**

**REGISTRATION NO. ⇒ 19BCE10071**

## **Data mining and Analysis**

**Data Mining** is the process of extracting important patterns from large datasets.

Simply, it is a process that is used to turn raw data into meaningful data. It is

otherwise also called Knowledge Discovery in Databases (KDD). The improvement

in computing prowess has allowed data mining to become streamlined and

mainstream. It helps the organizations build more innovative strategies, increase

sales, generate revenue, and grow a business by cost reduction.

Eg - The best example of a data mining application is in the E-commerce sector,

where websites display options for those who purchased and viewed the specific

product.

**Data Analysis** is the process of analyzing and organizing raw data to determine useful

information and decisions. It is a superset of Data Mining, which includes removing,

cleaning, changing, demonstrating the data to reveal significant and valuable insights.

Eg. - The best example of data analysis is the study of the census.

Additionally, Data Mining and Data Analytics are different words but they indeed are very similar. Data mining is a step in the process of data analytics. Data Analytics is the

umbrella that deals with every step in the pipeline of any data-driven model.

Before we jump into the Core Tasks, Applications & Algorithms of both the terms, let’s look

into some identities through which we can find some differences between both terms.

* Data mining is catering the data collection and deriving crude but essential

insights. Data analytics then uses the data and crude hypothesis to build upon

that and create a model based on the data.

* Data mining shines its brightest when the data in question is well structured.

Meanwhile, data analysis can be performed on any data; it would still be able to

derive meaningful insights that could help in propelling the corporation to even

greater heights.

* Data mining is tasked to accomplish the main job to make the data that is being

used more usable. Whereas, data analysis is used to hypothesize and, in the

end, culminate itself in providing valuable information to help in business

decisions.

* 4. Data mining does not need any bias or any notions which are instilled before

tackling the data. Whereas, data analysis is majorly used for hypothesis

testing.

* 5. Data mining uses scientific and mathematical models and methods to identify

patterns or trends in the data that is being mined. On the other hand, data

analysis is employed to task with business analytics problems and derive

analytical models.

* 6. Data mining usually does not need any visualizations, bar charts, graphs, GIPs,

etc., whereas these visualizations /are the bread and butter of data analysis.

Without a good representation of the data in question, all the efforts which are

put into the analysis of the data would not come to fruition.

**Core Tasks**

**Data Mining**: There are several data mining tasks such as classification, prediction,

time-series analysis, association rule learning, clustering, summarization, etc. All

these tasks are either predictive data mining tasks or descriptive data mining tasks.

A data mining system can execute one or more of the above-specified tasks as part

of data mining.

**Data Analytics**: The tasks in Data analytics depend on the data-driven

decision-making practices adopted by the organization. Some of the

general tasks are mentioned below:

1. Designing and maintaining data systems and databases

2. Mining data from primary and secondary sources

3. Using statistical tools to interpret data sets, paying particular attention

to trends and patterns could be valuable for diagnostic and predictive

analytics efforts.

Generally, there exist 2 methods through which tasks can be

performed: Qualitative research and Quantitative research.

**Application**

**Data Mining:** There exist many applications of data mining, but some

major ones are mentioned below:

1. Financial Analysis

2. Telecommunication Industry

3. Intrusion Detection

4. Retail Industry

5. Higher Education

6. Energy Industry

7. Spatial Data Mining

8. Biological Data Analysis

9. Other Scientific Applications

10. Manufacturing Engineering

11. Criminal Investigation

12. Counter-Terrorism

**Data Analytics:** Some of the major applications are mentioned below:

1. Transportation - strategies to plan alternative routes, reduce

congestions and traffics, optimize the buyer’s experience in the travels.

2. Logistics and Delivery - best shipping routes, approximate delivery

times, real-time status of goods.

3. Web Search or Internet Web Results - The searched data is

considered as a keyword and all the related pieces of information are

presented in a sorted manner that one can easily understand.

4. Manufacturing - prediction analysis, regression analysis, budgeting,

etc.

5. Security - Security Analytics, identify danger before it gets an

opportunity to affect your framework.

6. Education - adaptive learning, innovations, adaptive content, etc.

7. Healthcare - channel enormous measures of information in

seconds to discover treatment choices or answers for various

illnesses.

8. Military - augmented reality and psychological science

Algorithms

**Data Mining:** There are many algorithms but we’ll mention a few.

1. C4.5 Algorithm

2. K-mean Algorithm

3. Support Vector Machines

4. Apriori Algorithm

5. Expectation-Maximization Algorithm

6. PageRank Algorithm

7. Adaboost Algorithm

8. KNN Algorithm

9. Naive Bayes Algorithm

10. CART Algorithm

**Data Analytics:** For data analytics, first we’ll see the types of analytics -

1. Descriptive analytics examines what happened in the past: Monthly

revenue, quarterly sales, yearly website traffic, and so on. These

types of findings allow an organization to spot trends.

2. Diagnostic analytics considers why something happened by comparing

descriptive data sets to identify dependencies and patterns. This helps an

organization

determine the cause of a positive or negative outcome.

3. Predictive analytics seeks to determine likely outcomes by detecting

tendencies in descriptive and diagnostic analyses. This allows an

organization to take proactive action—like reaching out to a customer

who is unlikely to renew a contract, for example.

4. Prescriptive analytics attempts to identify what business action to take.

While this type of analysis brings significant value in the ability to

address potential problems or stay ahead of industry trends, it often

requires the use of complex algorithms and advanced technology such

as machine learning.

As data analytics is a superset of data mining, hence to perform the type of

analysis on data, we use the same types of algorithms as in data mining. The

best 5 algorithms in analytics for big data are Linear Regression, Logistics

Regression, CART, KNN, and K-Means.

References

* Top 10 Most Common Data Mining Algorithms You Should Know | upgrade a blog
* Top 8 Applications Of Data Analytics to look out for in 2021 (jigsawacademy.com)
* 5 Advanced Analytics Algorithms for Your Big Data Initiatives | Transforming Data with Intelligence (tdwi.org)
* Data Mining Tasks | Data Mining tutorial by Wideskills
* 13 Really Cool Quotes About Data | The TIBCO Blog