410253(C): Project Work

**Fourth Year Computer Engineering**

**Year 2022 - 2023**

Group/Project ID: **18**

Team Members:

|  |  |
| --- | --- |
| 1. **SONAWANE YOGESH SURYABHAN** | **( B191014275 )** |
| 1. **BARDE VAISHNAVI ANIL** | **( B191014207 )** |
| 1. **SANSARE JAYSHREE SANJAY** | **( B191014263 )** |
| 1. **PAWAR MAHESH POPATRAO** | **( B191014252 )** |

**Mini Project**

**Subject : Business Intelligence**

**Subject Teacher:**

**PROF. GADE. N. B.**

**Ahmednagar Jilha Maratha Vidya Prasarak Samaj’s**

**Shri Chhatrapati Shivaji Maharaj College of Engineering,**

**Nepti, Ahmednagar**

DEPARTMENT OF COMPUTER ENGINEERING

SCSMCOE, Department of Computer Engineering 2022-23

**SHRI CHHATRAPATI SHIVAJI MAHARAJ COLLEGE OF ENGINEERING**

**NEPTI, AHMEDNAGAR, (M.S.) INDIA- 414 005**

**DEPARTMENT OF COMPUTER ENGINEERING**



**CERTIFICATE**

This is to certify that the seminar entitled

**“****BUSINESS INTELLIGENCE MINI PROJECT REPORT”**

Submitted by

**Mr. Sonawane Yogesh Suryabhan**

SEAT NUMBER. -B191014275

This is to certify that **Mr Sonawane Yogesh Suryabhan** has successfully completed her seminar work on **“BUSINESS INTELLIGENCE MINI PROJECT REPORT”** at **SHRI CHHATRAPATI SHIVAJI MAHARAJ COLLEGE OF ENGINEERING NEPTI, AHMEDNAGAR** in partial fulfilment of the Graduate Degree course in **Fourth Year of Computer**, in Academic Year 2022-23 Semester-II as prescribed by the **“Savitribai Phule Pune University”**

|  |  |  |
| --- | --- | --- |
| Prof. Sisodiya Y.A.  Head Of Department |  | PROF. GADE N.S.  Subject Teacher |
| Department of Computer Engineering, SCSMCOE |  | Department of Computer Engineering, SCSMCOE |
| Place: Ahmednagar |  | Date: / / |

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr.**  **No.** | **Description** | **Page No.** |
| **1** | **Problem statement** | **1** |
| **2** | **OBJECTIVES & outcomes** | **1** |
| **4** | **HARDWARE & SOFTWARE REQUIREMENT** | **2** |
| **5** | **INTRODUCTION** | **3-6** |
| **7** | **CONCLUSION** | **7** |

1. **PROBLEM STATEMENT**

Each group of 4 Students (max) assigned one case study for this; A BI report must be prepared outlining the following steps :

a) Problem definition, identifying which data mining task is needed.

b) Identify and use a standard data mining dataset available for the problem.

1. **Objectives and Outcomes:**

**Objectives –**

* To introduce the concepts and components of Business Intelligence (BI)
* To develop problem solving ability
* To Organize, sustain and report on a substantial piece of team work over a period of several months
* To Evaluate alternative approaches, and justify the use of selected tools and methods,
* To Reflect upon the experience gained and lessons learned,
* To Consider relevant social, ethical and legal issues,
* To find information for yourself from appropriate sources such as manuals, books, research journals and from other sources, and in turn increase analytical skills.
* To Work in TEAM and learn professionalism.

**Outcomes -**

* Use tools and techniques in the area of software development to build mini projects
* Apply basic principles of elective subjects to problem solving and modelling.
* Analyse alternative approaches, apply and use most appropriate one for feasible solution.
* Write precise reports and technical documents in a nutshell.
* Participate effectively in multi-disciplinary and heterogeneous teams exhibiting team work, Inter-personal relationships, conflict management and leadership quality

1. **SOFTWARE REQUIREMENTS SPECIFICATIONS**

**System configurations**

The software requirement specification can produce at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by established a complete information description, a detailed functional description, a representation of system behavior, and indication of performance and design constrain, appropriate validate criteria, and other information pertinent to requirements.

**Software Requirements:**

* Operating system : windows10
* Coding Language : Java
* Software : PowerBI

**Hardware Requirement:**

1. System : INTEL i5 3 rd genPC
2. Hard Disk : 500GB
3. Ram : 4GB
4. **Introduction**

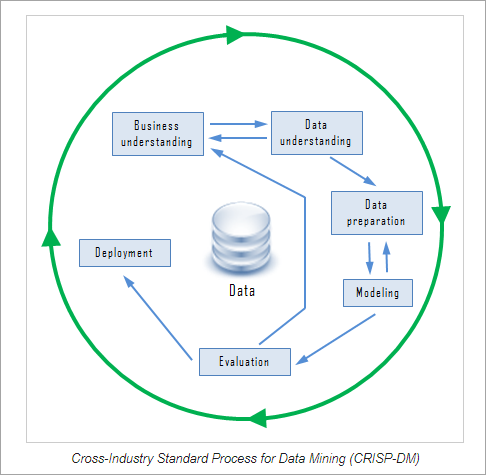
**Title: identifying which data mining task is needed.**

* 1. **DEFININTION:**

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools enable enterprises to predict future trends and make more-informed business decisions.

* 1. **IMPORTANCE**

Data mining is a crucial component of successful analytics initiatives in organizations. The information it generates can be used in business intelligence (BI) and advanced analytics applications that involve analysis of historical data, as well as real-time analytics applications that examine streaming data as it's created or collected.

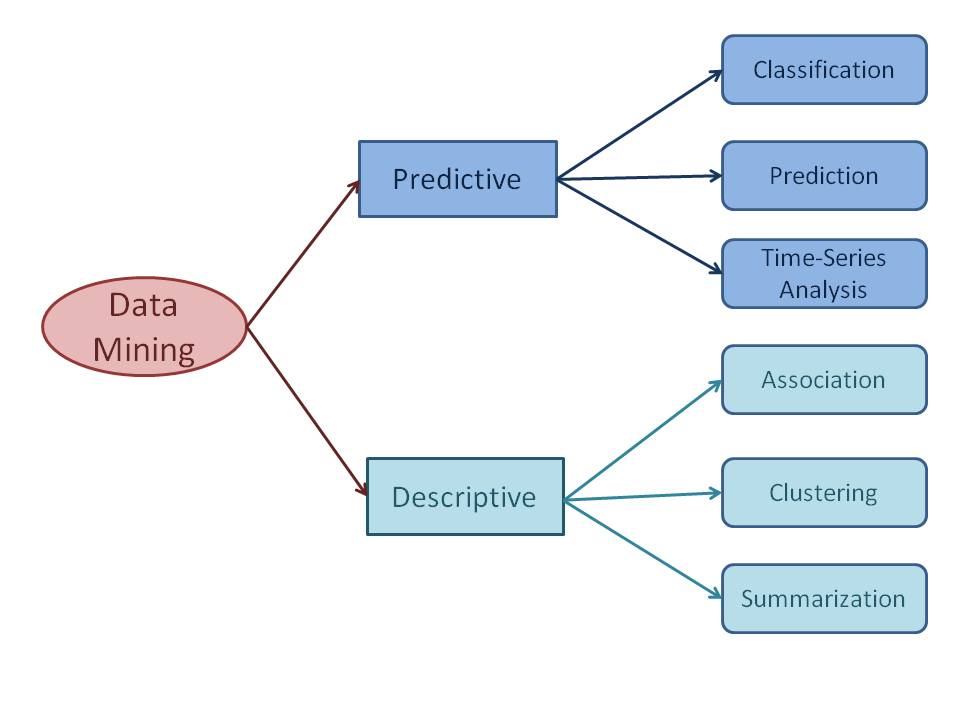


* 1. **Introduction to Data Mining Tasks**

The data mining tasks can be classified generally into two types based on what a specific task tries to achieve. Those two categories are descriptive tasks and predictive tasks. The descriptive data mining tasks characterize the general properties of data whereas predictive data mining tasks perform inference on the available data set to predict how a new data set will behave.

* + 1. **Different Data Mining Tasks**

There are a number of data mining tasks such as classification, prediction, time-series analysis, association, 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.



Predictive data mining tasks come up with a model from the available data set that is helpful in predicting unknown or future values of another data set of interest. A medical practitioner trying to diagnose a disease based on the medical test results of a patient can be considered as a predictive data mining task. Descriptive data mining tasks usually finds data describing patterns and comes up with new, significant information from the available data set. A retailer trying to identify products that are purchased together can be considered as a descriptive data mining task.

**a) Classification**

Classification derives a model to determine the class of an object based on its attributes. A collection of records will be available, each record with a set of attributes. One of the attributes will be class attribute and the goal of classification task is assigning a class attribute to new set of records as accurately as possible.

Classification can be used in direct marketing, that is to reduce marketing costs by targeting a set of customers who are likely to buy a new product. Using the available data, it is possible to know which customers purchased similar products and who did not purchase in the past. Hence, {purchase, don’t purchase} decision forms the class attribute in this case. Once the class attribute is assigned, demographic and lifestyle information of customers who purchased similar products can be collected and promotion mails can be sent to them directly.

**b) Prediction**

Prediction task predicts the possible values of missing or future data. Prediction involves developing a model based on the available data and this model is used in predicting future values of a new data set of interest. For example, a model can predict the income of an employee based on education, experience and other demographic factors like place of stay, gender etc. Also prediction analysis is used in different areas including medical diagnosis, fraud detection etc.

**c) Time - Series Analysis**

Time series is a sequence of events where the next event is determined by one or more of the preceding events. Time series reflects the process being measured and there are certain components that affect the behavior of a process. Time series analysis includes methods to analyze time-series data in order to extract useful patterns, trends, rules and statistics. Stock market prediction is an important application of time- series analysis.

**d) Association**

Association discovers the association or connection among a set of items. Association identifies the relationships between objects. Association analysis is used for commodity management, advertising, catalog design, direct marketing etc. A retailer can identify the products that normally customers purchase together or even find the customers who respond to the promotion of same kind of products. If a retailer finds that beer and nappy are bought together mostly, he can put nappies on sale to promote the sale of beer.

**e) Clustering**

Clustering is used to identify data objects that are similar to one another. The similarity can be decided based on a number of factors like purchase behavior, responsiveness to certain actions, geographical locations and so on. For example, an insurance company can cluster its customers based on age, residence, income etc. This group information will be helpful to understand the customers better and hence provide better customized services.

**f) Summarization**

Summarization is the generalization of data. A set of relevant data is summarized which result in a smaller set that gives aggregated information of the data. For example, the shopping done by a customer can be summarized into total products, total spending, offers used, etc. Such high level summarized information can be useful for sales or customer relationship team for detailed customer and purchase behavior analysis. Data can be summarized in different abstraction levels and from different angles.

* 1. **Industry examples of data mining**

Here's how organizations in some industries use data minin**g** as part of analytics applications:

* **Retail.**

Online retailers mine customer data and internet clickstream records to help them target marketing campaigns, ads and promotional offers to individual shoppers. Data mining and predictive modeling also power the recommendation engines that suggest possible purchases to website visitors, as well as inventory and supply chain management activities.

* **Financial services.**

Banks and credit card companies use data mining tools to build financial risk models, detect fraudulent transactions and vet loan and credit applications. Data mining also plays a key role in marketing and in identifying potential upselling opportunities with existing customers.

* **Insurance.**

Insurers rely on data mining to aid in pricing insurance policies and deciding whether to approve policy applications, including risk modeling and management for prospective customers.

* **Manufacturing.**

Data mining applications for manufacturers include efforts to improve uptime and operational efficiency in production plants, supply chain performance and product safety.

* **Entertainment.**

Streaming services do data mining to analyze what users are watching or listening to and to make personalized recommendations based on people's viewing and listening habits.

* **Healthcare.**

Data mining helps doctors diagnose medical conditions, treat patients and analyze X-rays and other medical imaging results. Medical research also depends heavily on data mining, machine learning and other forms of analytics.

1. **CONCLUSION**

Data Mining is an iterative process where the mining process can be refined, and new data can be integrated to get more efficient results. Data Mining meets the requirement of effective, scalable and flexible data analysis.

It can be considered as a natural evaluation of information technology. As a knowledge discovery process, Data preparation and data mining tasks complete the data mining process.

Data mining processes can be performed on any kind of data such as database data and advanced databases such as time series etc. The data mining process comes with its own challenges as well