UNIVERSITY O F PETROLEUM AND ENERGY STUDIES

UPES
UNIVERSITY OF THE FUTURE

BACHELOR OF TECHNOLOGY

Computer Science Engineering

(specialization in Artificial Intelligence And Machine Learning)

A DATA MINING FRAMEWORK TO ANALYZE ROAD ACCIDENT DATA

High Level Design

25/07/2022

Version No.	1.00
Authorized by	PIYUSH MALVIYA PRIYAL GUPTA SAKSHAM GARG RAHUL DHANOLA

Industry Mentor:
Sumit Shukla

Guided By:
Sujoy Chatterjee

MEMBERS

S. NO.	NAME	ROLL NO.	BATCH	PHONE NO.	EMAIL ADDRESS
1.	Piyush Malviya	R177219128	B4	6261622231	500075183@stu.upes.ac.in
2.	Priyal Gupta	R177219136	B6	9415209873	500076110@stu.upes.ac.in
3.	Rahul Dhanola	R177219139	B4	6397227162	500075154@stu.upes.ac.in
4.	Saksham Garg	R177219154	B5	9996619712	500076000@stu.upes.ac.in

Table of Contents

Figures Used

Introduction

1.1 Scope of the document

This document outlines the high level functional design of data mining framework analyzing road accidents. It highlights the high level Use cases in analyzing and predicting prone areas of road accidents and their causes and design of components. It serves as an input to the low level design documents that would further elaborate on the application design.

This document outlines the high level design of the machine learning model that is used to design a data framework to analyze the road accidents in India. This document is a reference design document for developers/implementers so that they can develop systems with minimal effort.

Road accident analysis aims to investigate the main factors that characterize an accident to understand patterns or behaviors and, consequently, to identify the appropriate countermeasures to adopt to avoid the accident.

Researchers have aimed to find out which data mining techniques are most suitable to analyze road accidents, to identify the most significant causes and the most recurrent patterns of road accidents.

1.2 Intended Audience

- Traffic engineers and government agencies.
- Transportation department to analyze and run instructional recommendation systems.
- Development authorities for constructing effective relevance.
- Firms which deal with accidents and preventions.

1.3 System overview

This system provides a well analyzed data mining framework which indicates the accidents and their causes according to the authentic data obtained from data.gov.in .In this we will use different data mining algorithms to analyze the data. Different algorithms are applied to group the accident locations into clusters and mining techniques are used to characterize the locations.
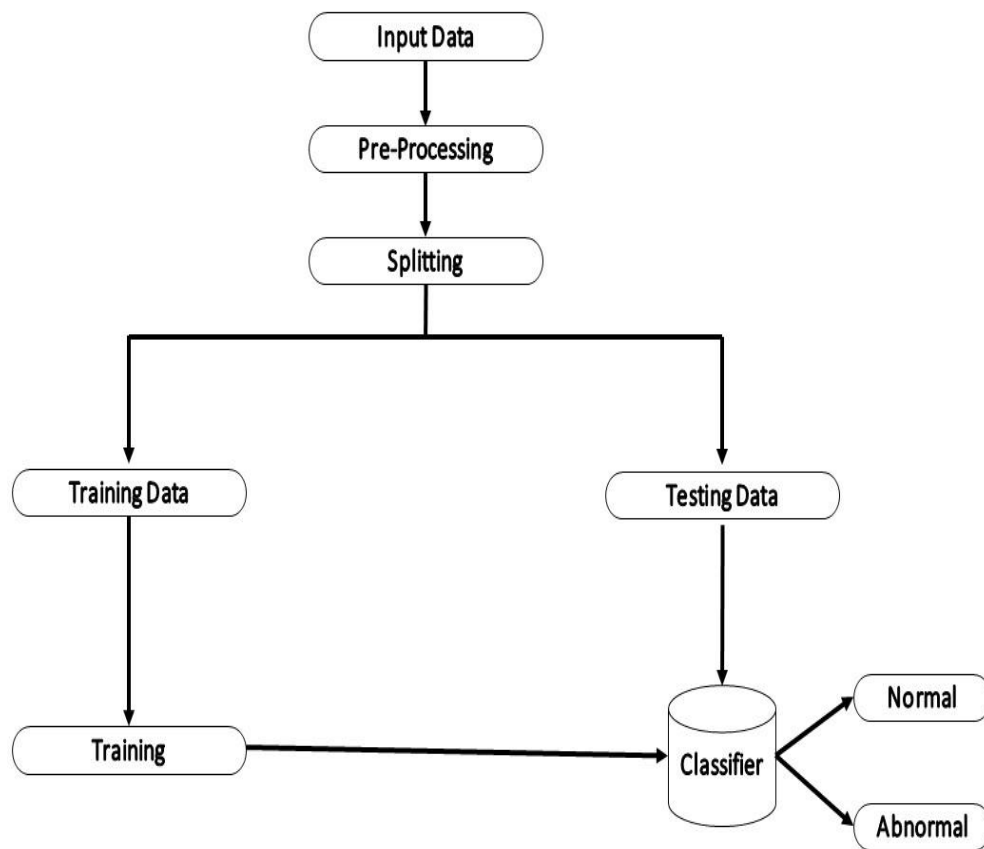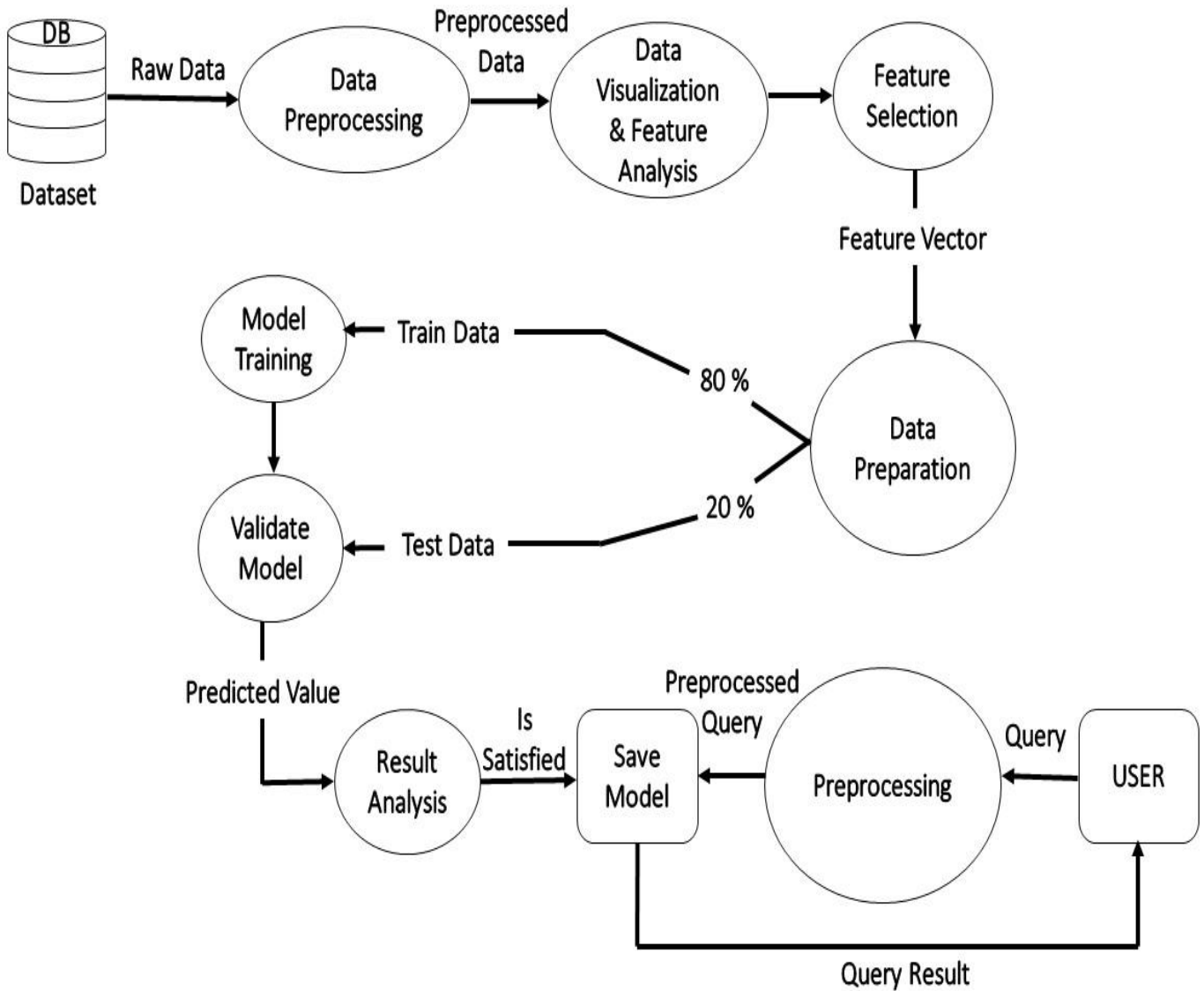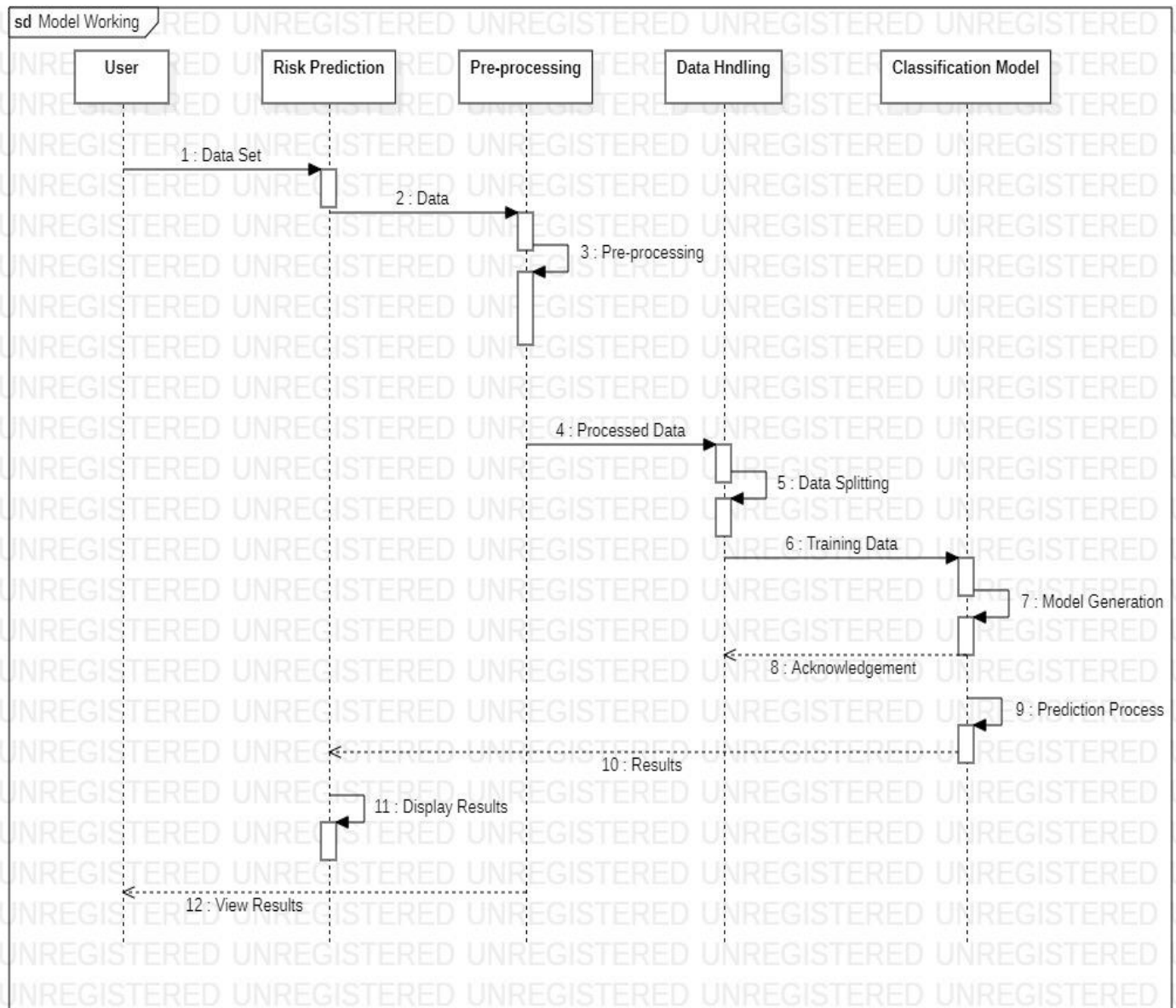

Fig1: System Block Diagram

Fig2: Data Flow of System

I

2.High Level System Design

2.1 Sequence Diagram:

sd Model Working

| User | Risk Prediction | Pre-processing | Data Hndling | Classification Model |

1 : Data Set

2 : Data

3 : Pre-processing

4 : Processed Data

5 : Data Splitting

6 : Training Data

7 : Model Generation

8 : Acknowledgement

9 : Prediction Process

10 : Results

11 : Display Results

12 : View Results

2.2System Configuration

This project can run on commodity hardware. First part is the training phase which takes 10-15 mins of time and the second part is the testing part which only takes a few seconds to perform classification and calculate accuracy.

Hardware Requirements

Table 1: Table for Hardware Requirements

Type	Name
System Architecture	32 − bit or 64 − bit
Memory	8 GB 2400 MHz DDR4
Storage	500 GB SATA3 2.4 HDD
Central Processing Unit (CPU)	Intel or AMD - 2 GHz or Faster
Graphical Processor Unit (GPU)	2 GB Nvidia Graphic Processor

Software Requirements

Table 2: Table for SoftwareRequirements

Type	Name
Programming Language with Version	Python 3.5 & above
Operating System	Windows 7 & above. Linux based OS Mac OS

3.Details of other frameworks being used

Session Management:

In this we will use different data mining algorithms to analyze the data. Different algorithms are applied to group the accident locations into clusters and mining techniques are used to characterize the locations. Most state of the traffic management and information systems focus on data analysis. Python, and Jupyter notebooks are mainly used. Since Python has a large number of libraries and packages, it has a very

large ecosystem. Python is used both in data scraping and in developing the server. Jupyter notebook is an open-source and web based interactive environment for making notebook documents. The primary Jupyter online application and Jupyter python web server are the substances required for making a notebook.

4.Unit Testing

We will develop and train the machine learning model from scratch and subsequently a testing framework will also be made for the classification module. We will use many classification datasets to test the model .We will also document the performance and accuracy of our model.

5.References

[1] A. Priyanka and K. Sathiyakumari, "A comparative study of classification algorithm using accident data", International Journal of Computer Science & Engineering Technology (IJCSET)

[2] Karlaftis M, Tarko A. Heterogeneity considerations in accident modeling. Accid Anal Prev. 1998;30(4):425–33.

[3] Article(online) Available:
https://www.hindustantimes.com/mumbai-news/india-had-most-deaths-in-road-accidents-in-2019-report/story-pikRXxsS4hptNVvf6J2g9O.html

[4] Road Traffic Accidents in India: Issues and Challenges, Sanjay Kumar Singh
https://www.sciencedirect.com/science/article/pii/S2352146517307913

[5] Ng KS, Hung WT, Wong WG. An algorithm for assessing the risk of traffic accidents. J Saf Res. 2002;33:387–410.

[6] T. Dipo, I Akomolafe and Akinbola Olutayo, "Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways", American Journal of Database Theory and Application 2012

[7] Article by S. Nagendra Babu "A Data Mining Framework to Analyze Road Accident Data using Map Reduce Methods CCMF and TCAMP Algorithms"