REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Accident Severity Detection and Prediction Using Machine Learning Techniques

SHARON JOSEPH

SRN: R19DM053

Date: 20-08-2022

PGDM/MBA in Business Analytics

Capstone Project Presentation
Year: I

race.reva.edu.in

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Agenda

01 Introduction
Background | Current status | Why this study

02 Literature Review
Seminal works | Summary | Research Gap

03 Problem Statement
Business Problem | Analytics Solution

04 Project Objectives
Primary & Secondary Objectives | Expected Outcome

05 Project Methodology
Conceptual Framework | Research Design

06 Business Understanding
Business Context | Monetary Impact

07 Data Understanding
Data Collection | Variables

08 Data Preparation
Pre-processing | Process | Techniques

09 Descriptive Analytics
Univariate | Bivariate | Hypothesis

10 Modeling
Machine Learning | Model Evaluation | Insights

11 Model Deployment
Applications | Demo

12 Suggestions and Conclusions
Insights | Next Step | Future Scope

13 Annexure
References | Publications | Plagiarism Score

Introduction

Background:

There has been a drastic increase in the economic activities and growing human needs, leading to the expansion of travel and transportation. Therefore, the increase in vehicles and traffic movement has led to road accidents, influenced public health and the country's economy.

Why this study :

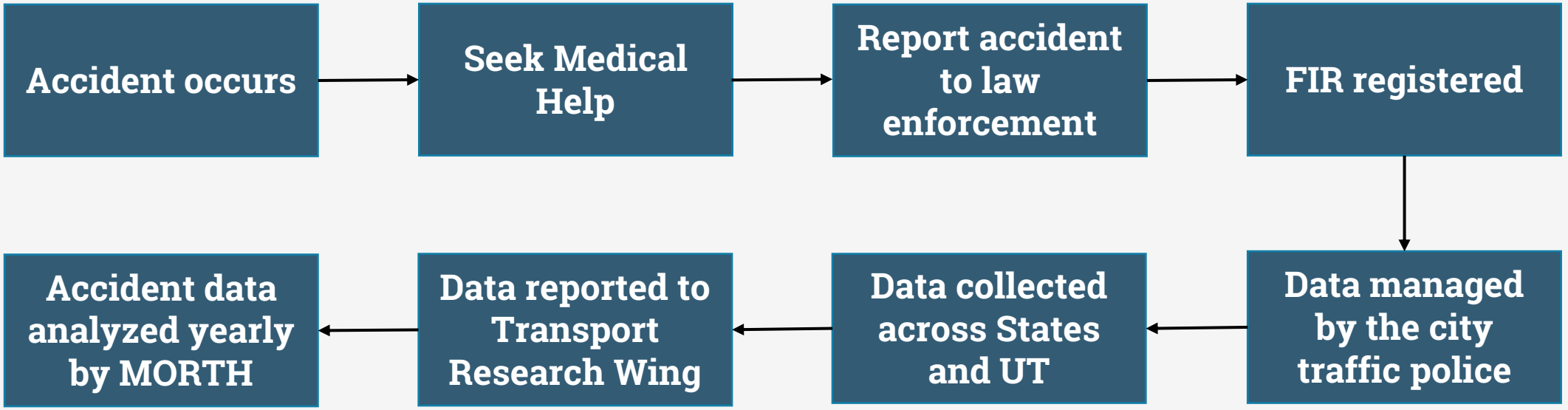
Working for an Automobile Industry (VOLVO), wanted to build an analytical solution, which would help analyze accident-causal factors and predict accidents' severity, thereby helping our community towards safer travel on Indian roads.

1.3M Accidents Worldwide

3.6L Accidents in India

1744 Accidents in Bangalore

How is traffic accident data managed today?



Abbreviations:

FIR - First Information Report

UT – Union Territory

MORTH – Ministry of Road Transport and highways

Literature Review

Paper Title	Author	Year	Summary	Research Gap
Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia	Tibebe Beshah, Shawndra Hill	2010	Data mining technologies were applied to link recorded road characteristics to accident severity in Ethiopia and developed a set of rules that could be used by the Ethiopian Traffic Agency to improve safety.	Focussed on Ethiopia's Road-related factors only.
Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques	S. Shanthi, R.Geetha Ramani	2012	Road traffic accidents are classified based on injury severity using classification algorithms such as C4.5, CR-T, ID3, CS-CRT, CS-MC4, Naïve Bayes and Random Tree. In this research work we used the road accident training dataset which was obtained from the Fatality Analysis Reporting System (FARS), provided by the University of Alabama's Critical Analysis Reporting Environment (CARE) system.	Focused on the factors such as Manner of Collision, Driver Seating Position, Harmful Event, Protection System, Driver Age and Drug Involvement.
Traffic AccidentAnalysis UsingDecision Trees and Neural Networks	OlutayoV.A, EludireA.A	2014	This work employedArtificial Neural Networksand Decision Trees data analysis techniques to discover new knowledge from historical data about accidents in one of Nigeria's busiest roads in order to reduce carnage on our highways.	Studies the causal factors of accidents in Nigeria which mainly involved tyre burst, broken shaft and loss of vehicle control.

Literature Review

Paper Title	Author	Year	Summary	Research Gap
Spatial Decision Tree for Accident Data Analysis	J. M. Manasa, Shrutilipi Bhattacharjee, Soumya K. Ghosh and Sudeshna Mitra	2015	This paper aims at categorizing and analyzing the accident data and drawing some meaningful inferences, that are implicit to the data. A spatial decision tree-based approach has been used and implemented to draw some useful conclusions. The experimentation has been carried out on the accident dataset, collected from the National Highway (NH6) connecting Kharagpur and Kolkata, India.	Prime focus is to identify accident hot spots using spacial decision trees.
Analyzing Factors, Construction of Dataset, Estimating importance of factor and generation of association rules for Indian road Accident	Suwarna Gothane, Dr. M. V. Sarode	2016	In this paper, the factors influencing the accidents are identified and the factor which is most accident prone is studied with Info Gain Attribute Evaluator function using the WEKA tool.	Here the attribute's importance is evaluated based on info gain attribute approach.
Data mining approach to analyse the road accidents in India	Ayushi Jain, Garima Ahuja, Anuranjana, Deepti Mehrotra	2016	The purpose of this paper is to employ data mining to assist in the development of a prototype that not only provides a broad classification of information by combining related components with each other in order to identify the accident-prone regions of the country based on different accident-causing variables but also clearly signifies the relationship between these variables and casualties	Cluster analysis is used to determine the accident-prone states and territories of India to conclude the dominant factor, backing the accidents
Analysis of Datamining Techniquefor Traffic Accident Severity Problem: A Review	Meenu Gupta, Vijender Kumar Solanki, Vijay Kumar Singh	2017	A review of many researches in traffic safety area has been done and the road accident severity survey using data mining approach is being studied to reduce accidents leading to fatality of life.	No findings, Just a review of different research methods in this area.
Traffic accident's severity prediction: A deep-learning approach-based CNN network	Ming Zheng, Tong Li, Rui Zhu, Jing Chen, Zifei Ma, Mingjing Tang, Cui and Wang	2019	To promote the prediction accuracy,a novel traffic accident's severity prediction-convolutional neural network (TASP-CNN) model for traffic accident's severity prediction	There is a need to understand whether the TASP-CNN model can be applied to other areas, and whether it is more accurate than the existing models.

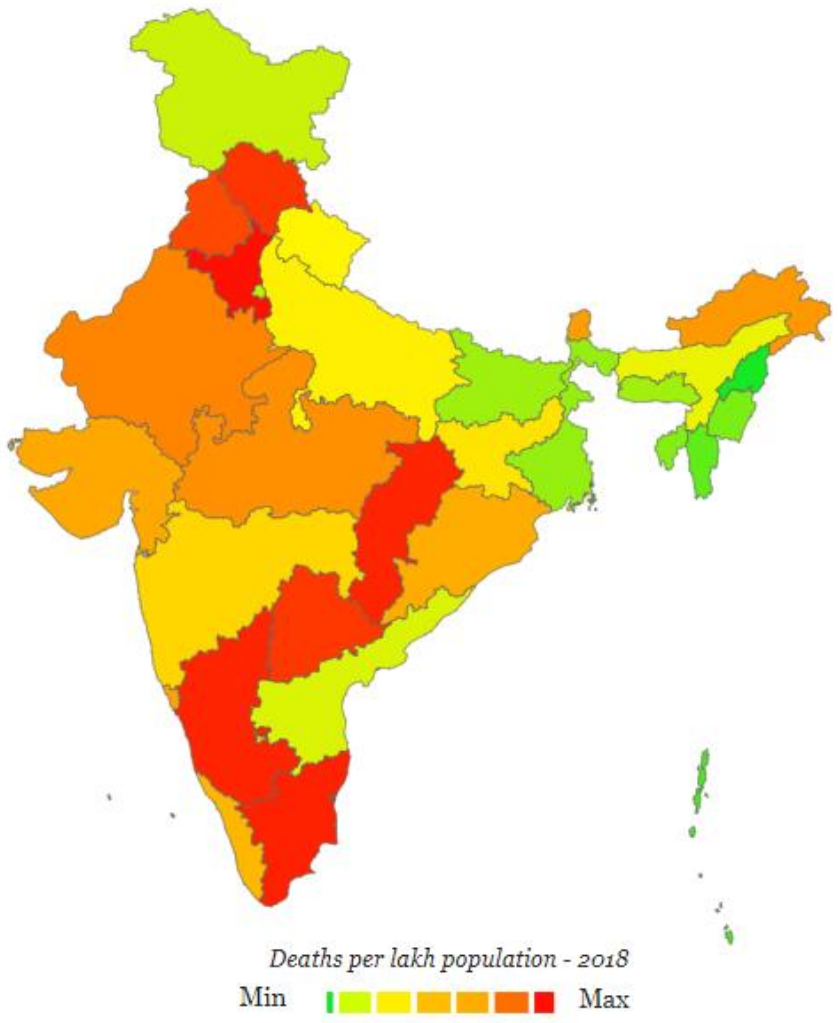
Literature Review

Paper Title	Author	Year	Summary	Research Gap
Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity	Rabia, Kenneth, Maha Reda, Abdulbaset	2019	Supervised machine learning algorithms, such as AdaBoost, Logistic Regression (LR), Naive Bayes (NB), and Random Forests (RF) are implemented on traffic accident data to predict the traffic accident severity	Studied on Michigan Traffic Data.
Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh	Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine	2019	This research paper has been done to analyze traffic accidents more deeply to determine the intensity of accidents by using machine learning approaches in Bangladesh.	Studied on Bangladesh Traffic Data.
Traffic accident prediction based on CNN model	Amani Thaduri, Vijayakumar Polepally, SwathyVodithala	2021	This paper proposes deep learning-based new road traffic accident prediction applying a "Convolutional Neural Network model" (CNN) and proves it to better than the traditions backpropagation method.	This prediction model relies entirely on the CNN to predict traffic accidents on highways
RFCNN: Traffic Accident Severity PredictionBased on Decision Level Fusion of Machineand Deep Learning Model	Mubariz, Umer, Sadiq, Abid, Saleem, Hamza, Bisogni	2021	This study presents an ensembleof machine learning and deep learning models by combining Random Forest and Convolutional NeuralNetwork called RFCNN for the prediction of road accident severity.	The data used in the analysis include accident records of the USA.
Prediction of Traffic Accidents Severity Based on Machine Learning and Multiclass Classification Model	Mateja Iveta, Aleksander Radovan, Branko Mihaljević	2021	As per this study, the data points were supplied by the UK government and were utilized as chronological data to train the models from January 2015 to December 2018.The prediction results show the correlation between weather conditions, daylight time, and traffic accident severity using multiclass classification.	The datasets used were related to road accidents in several countries in a period of a few years.

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Problem Statement:

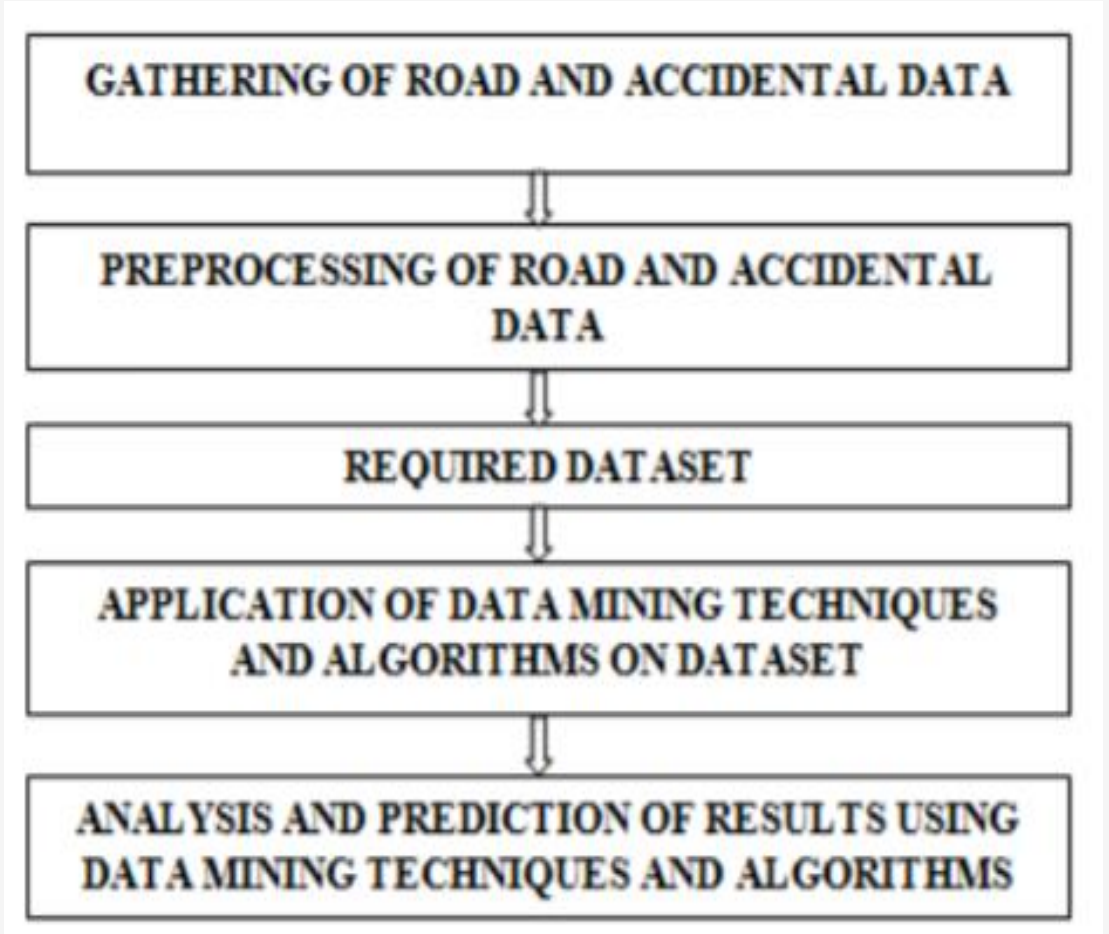
- The rising number of deaths and injuries as a result of traffic accidents is acknowledged to be a global phenomenon and traffic safety has been a serious concern since the start of the automobile age. The roads of India haven't reduced their contribution in the accident fatalities.
- The conventional method is manual, in which the data is manually tracked in a ledger system to analyse the information and based on the analysis, they take preventative measures to decrease the number of accidents.



Source: Road Accidents Dashboard by MORTH.

Proposed Solution

The **Proposed system** uses data mining to help identify the major factors responsible for road accidents and predict the severity of accidents that may occur, thus helping the transport authorities in improving safety standards for travel on Indian roads.



Project Objectives

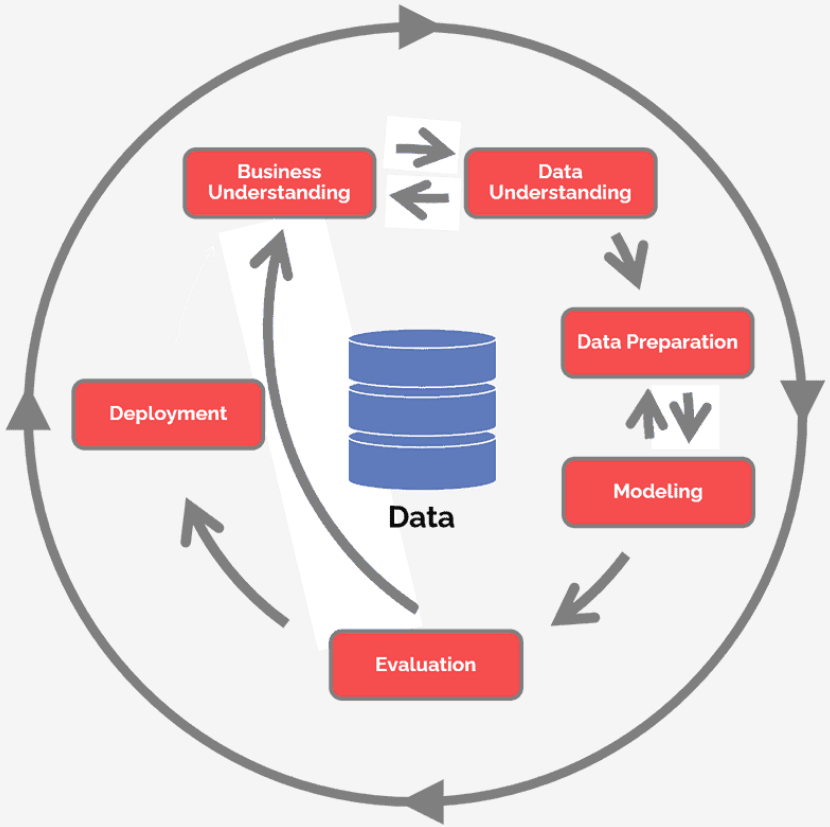
Two major objectives of this study are,

Identify the key features which contribute to traffic accidents.

Predict the class of accident severity into Fatal, Serious and Minor.

Project Methodology

CRISP-DM Framework : The Cross-Industry Standard Process for Data Mining methodology is used in the implementation of this project.



Source: CRISP-DM from
https://www.datascience-pm.com/crisp-dm-2/

Business Understanding: Understand the problems faced by common people, while traveling on Indian Roads.

Data Collection: Various Data like Weather Data, Accidents Data, Roads and Traffic data need to be collected.

Data Processing: Establish data connection, analyse and clean data, perform EDA.

Modelling: Apply Machine Learning algorithms to predict the accident severity.

Evaluation: Evaluate results and compare which algorithms give accurate results.

Deployment: Web Application to show the predicted severity of accidents based on various factors.

Business Understanding

The Problem addressed in this study is more of a **Community problem** than a business problem!

Accidents due to Road and Environment Factors.

Accidents due to Vehicular Factors.

Accidents due to Human Factors.

Untimely rescue operations

Law and Order

If the accident-causing factors are handled appropriately, the huge number of accidents can be avoided thereby saving Life, Time and Money

Challenges faced:

- There is no combined open-source data available for Indian traffic accidents along with the weather, road, vehicle and driver details.
- Traffic police maintains data locally in their systems which is collected district-wise , then state-wise and finally sent to MORTH to create yearly reports.
- These reports prepared by MORTH are not available for the recent years, last available data is for 2020.

Data Understanding

DATA COLLECTION:

- Data is collected from MORTH website from 2001 to 2020 to analyse the accidents that have occurred during the different times of the day and during the different months of the year

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

DATA DICTIONARY

These are the features considered in the dataset.

Variables	Description
Carriageway/Roadway hazards	This variable talks about any obstacles on the roadway. It could be animals or pedestrians, potholes, etc. Here we are generally classifying them as hazards present and none.
Light condition	This variable connotes the lighting during the accident. Variable represents Namely: Light present, Darkness
Day of the Week	Days through Sunday to Saturday
Special Conditions at the accident site	This includes malfunctioning of traffic lights, improper signboards, road maintenance issues, roadworks and fuel issues
Road Class	A-Principal Road, B-Secondary Road, C-Minor Road, Motorway and Unclassified Roads.
Junction Control	This field specifies how the junction is controlled.
Junction Details	Whether the accident occurred at a junction, crossroads, round-about, private driveway, and so on.
Road Surface condition	This variable shows the situation of the road during the accident. Variable represents Namely: dry, wet, frost, snow, flood
Road Type	This variable basically talks about the kind of roadway in which the accident has occurred. Like: Single Road, Double Road, One-way, Slip Road, U-turn.
Area Type	Whether Urban or Rural area.
Weather condition	This variable indicates the climate condition during the accident. Variable represents: Rain, Wind, Snow, Fog, Mist, Sunny, Fine weather
Speed	This variable talks about the speed of the vehicle at the time of the accident. In this dataset, it varies from 10-70 km/hr.
Accident time	This variable implies the time on which road traffic severity occurred throughout the day (24 hours) and whether day or night.
Severity	This variable is the target variable, and it represents three classes, namely: fatal, serious and minor injury.

Data Preparation

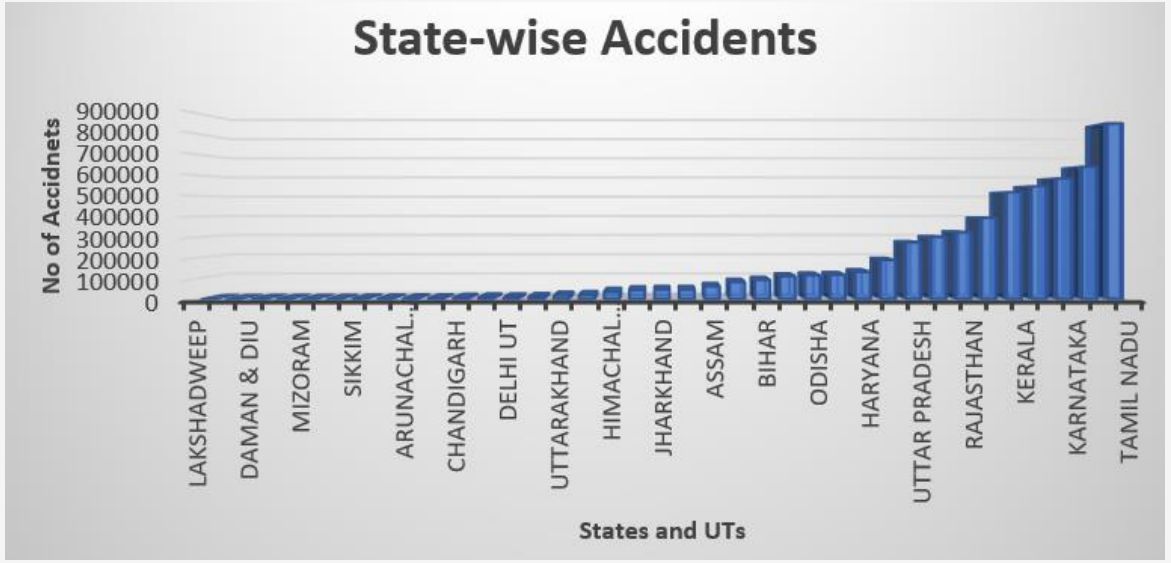
Data Cleaning

Encoding the data

Feature Engineering

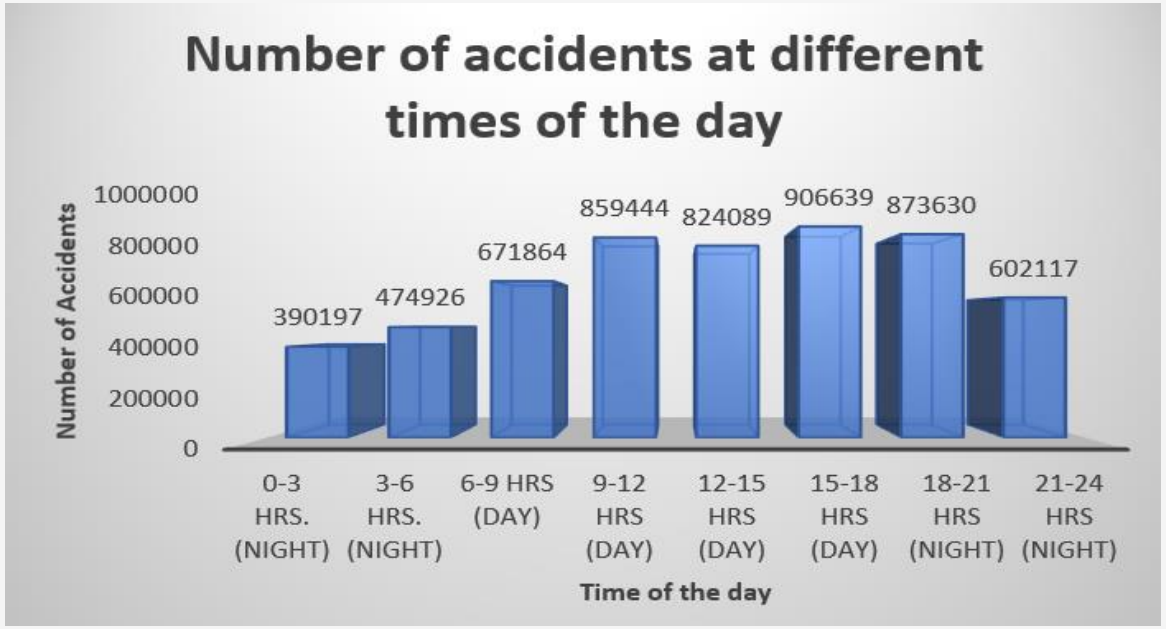
- Data has 1048575 rows and 14 columns
- Data is split such that train and test data should have prediction classes in equal ratio
- Data is split into a 75:25 ratio, where 75% of data is used to train the model and 25% of the dataset is used to measure the performance of the prediction model.
- This prepared data set will be the basis of the analysis.

Descriptive Analytics



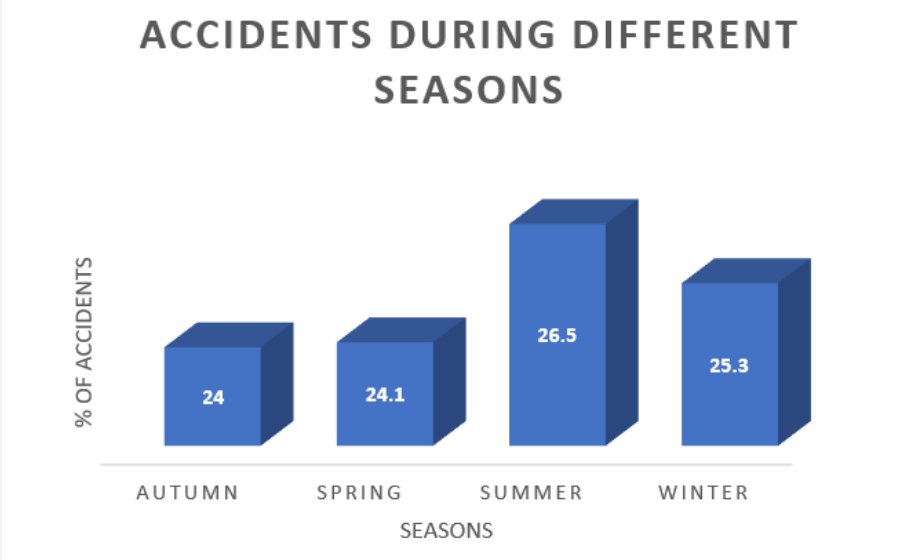
This chart shows that Tamil Nadu is the state with highest number of accidents across the country.

This chart shows that most accidents have occurred in the afternoon between 3:00pm and 6:00pm



Descriptive Analytics

This chart shows that highest number of accidents have happened in the month of May.



This chart shows that most accidents have during Summer season i.e.. in the months of March, April and May

Descriptive Analytics

Variables	Type	Maximum %
Road Class	A	45%
Carriageway Hazards	None	98%
Day	Friday	16%
Junction Control	Uncontrolled	48%
Junction Detail	Not at Junction	41%
Light Conditions	Daylight	73%
Road Surface	Dry	68%
Road Type	Single carriageway	74%
Special Conditions at the site	None	97%
Speed Limit	30	63%
Area	Urban	64%
Weather Conditions	No High Winds	79%
Accident severity	Minor	85%

Based on Frequency Analysis of the accident data, it is observed that most accidents from the dataset, have occurred in Road type A, which is on Main Roads, in Urban area where there is no junctions nor junction control, there are no special conditions or hazards on the roads and the weather looks good, speed is controlled, there is enough daylight and most accidents occurred on Friday. Therefore, the Accident severity of most accidents (around 85%) is **Minor** from the data that we have.

Modeling

Overview of the dataflow into Machine Learning Model

Data that has been cleaned and pre processed is modelled using multiple machine learning algorithms like the Gaussian Naïve Bayes and Random Forest classifier.

Of both these algorithms , the Gaussian Naïve bayes algorithm gave better results in terms of accuracy and the model performance. Hence the same has been used for the final prediction model.

Model Evaluation

The models are evaluated using the confusion matrix and the model performance is calculated using the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) values.

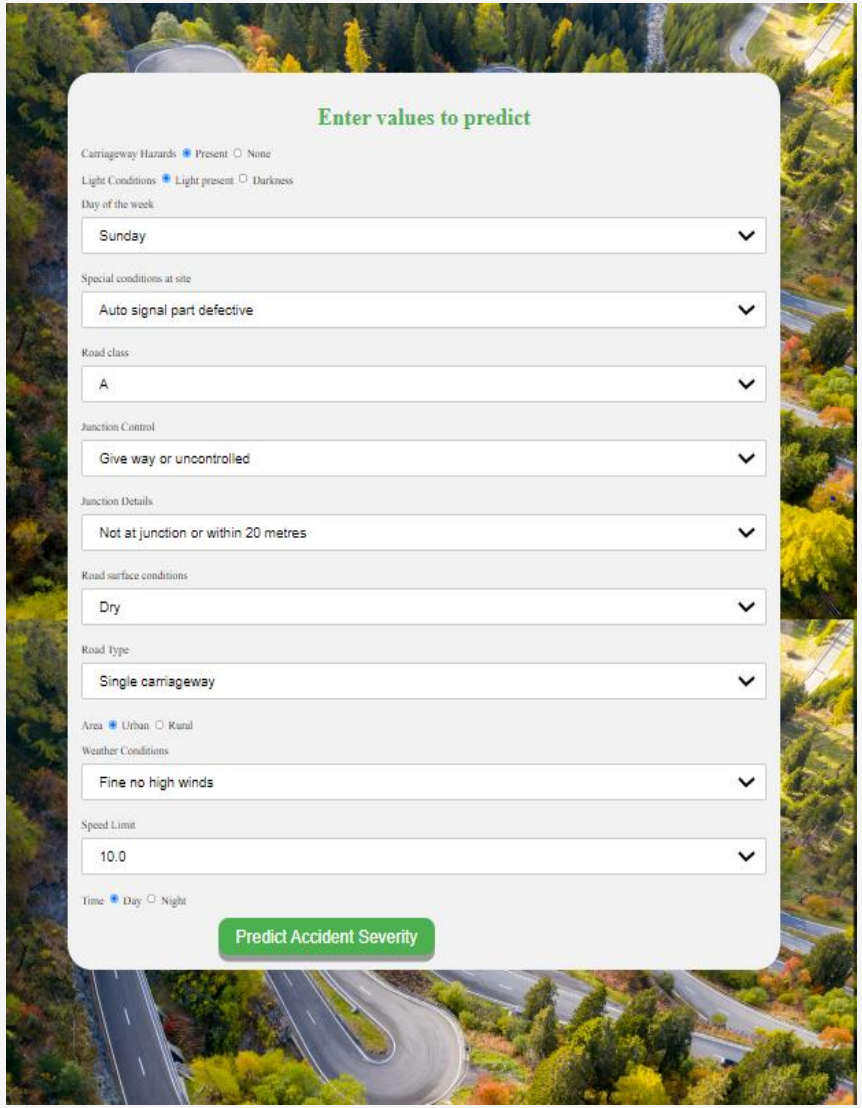
The Gaussian Naïve Bayes model has a higher accuracy score of 85% and an F1-score of 0.92 which is closer to 1, proving better performance.

Model Performance	Gaussian Naïve Bayes	Random Forest Tree
Accuracy	**0.85**	0.65
Precision	0.86	0.88
Recall	0.99	0.68
F-1 Score	**0.92**	0.77

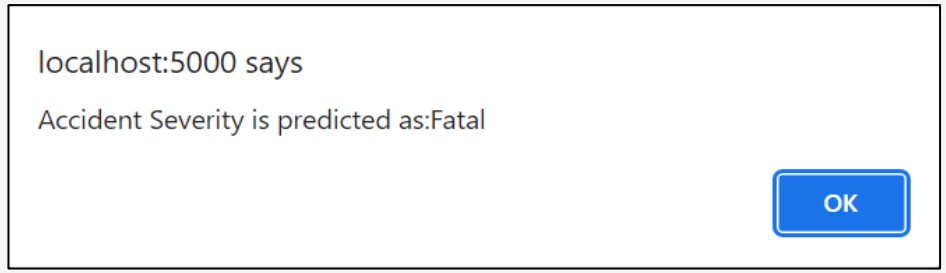
Performance Metrics

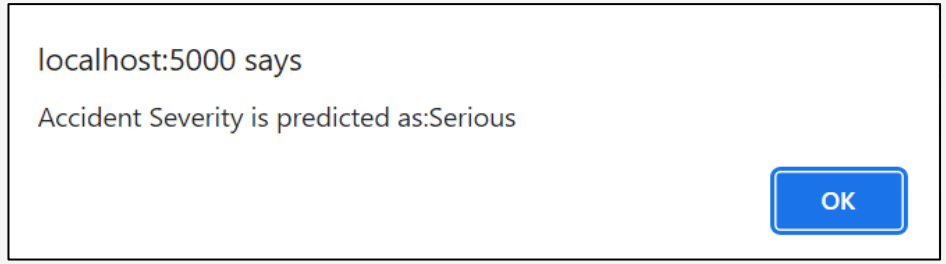
Model Deployment

The Python Flask framework is used to deploy the model on a web application. The prediction model is designed as a form-page in a web application, allowing for a more straightforward and simple method of foretelling the road accident intensity. The system can be installed in the end user domain as in the case of traffic police department.



After the form is filled, when the button called Predict Accident Severity is clicked, there is a window which displays the severity class predicted as minor, serious or fatal.





Results and Insights

Gaussian Naïve Bayes and Random Forest were the two algorithms that were explored during the study and the results showed that Naïve Bayes gave better accuracy of 85% and it is better and more optimal when lesser number attributes are used for analysis.

Based on the several accident causal factors, accidents severity is classified into Fatal, Serious and Minor.

Insights

- More Vacationers Hitting the Road could be the reason why most accidents happen in summer.
- The main reason for most accidents taking place between 3 PM and 6 PM is a result of huge rush of vehicles when people start returning to their homes after finishing work.
- Expansion in the road network, surge in motorization, rising population and huge number of accident hot spots is the reason why Tamil Nadu has highest number of accidents in India.

Conclusion and Future Work

Conclusion:

- In this project, data mining has helped in identifying India's accident-prone states and union territories. The main goal of this study is to evaluate and investigate accident-causal factors and predict the severity of the accidents that have occurred and that may occur in the future.
- Using data-mining methods can help in discovering how the time, date, road conditions and wind patterns are related to the various accidental severities caused. Machine learning algorithms like "Logistic Regression, Gaussian Naïve bayes, and Random Forest Classifier" are used to classify accident severity into Serious, Fatal and Minor accidents based on the several undermining factors identified during the data mining activity.
- The government is thus able to develop better road safety initiatives, indicate roads with signboards, notifying pedestrians and drivers of potential hazards, and design better road infrastructure.

Scope of future study:

- Vehicle and Driver information could be a good scope for further study to analyse the accident causal factors.
- Collect the precise accident location details and analyse for accident hot spots, so that appropriate measures could be taken by the traffic and safety department.

References

Bibliography | Webliography

Aman Kharwal. (2021). *Classification Report*. https://thecleverprogrammer.com/2021/07/07/classification-report-in-machine-learning/#:~:text=A%20classification%20report%20is%20a,this%20article%20is%20for%20you.

Beshah, T., & Hill, S. (2010). *Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia*.

Emhamed AlMamlook, R., Morgan Kwayu, K., Reda Alkasisbeh, M., & Ali Frefer, A. (2019). *Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity; Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity*.

Gothane, S., & Sarode, M. v. (2016). Analyzing Factors, Construction of Dataset, Estimating Importance of Factor, and Generation of Association Rules for Indian Road Accident. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, 15–18. https://doi.org/10.1109/IACC.2016.13

Gupta, M., Solanki, V. K., & Singh, V. K. (2017). Analysis of Datamining Technique for Traffic Accident Severity Problem: A Review. *Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering*, *10*, 197–199. https://doi.org/10.15439/2017r121

Iveta, M., Radovan, A., & Mihaljevic, B. (2021). Prediction of Traffic Accidents Severity Based on Machine Learning and Multiclass Classification Model. *2021 44th International Convention on Information, Communication and Electronic Technology, MIPRO 2021 - Proceedings*, 1701–1705. https://doi.org/10.23919/MIPRO52101.2021.9596664

Jain, A., Ahuja, G., Anuranjana, & Mehrotra, D. (2016). Data mining approach to analyse the road accidents in India. *2016 5th International Conference on Reliability, Infocom Technologies and Optimization, ICRITO 2016: Trends and Future Directions*, 175–179. https://doi.org/10.1109/ICRITO.2016.7784948

Jason Brownee. (2020). *gaussian classifier*. https://machinelearningmastery.com/gaussian-processes-for-classification-with-python/

Manasa, J. M., Bhattacharjee, S., Ghosh, S. K., & Mitra, S. (2015, February 9). Spatial decision tree for accident data analysis. *9th International Conference on Industrial and Information Systems, ICIIS 2014*. https://doi.org/10.1109/ICIINFS.2014.7036565

Manzoor, M., Umer, M., Sadiq, S., Ishaq, A., Ullah, S., Madni, H. A., & Bisogni, C. (2021). RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model. *IEEE Access*, *9*, 128359–128371. https://doi.org/10.1109/ACCESS.2021.3112546

Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, & Faria Nawrine. (2019). *Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh*.

Nick Hotz. (2022). *Data Science Process*. https://www.datascience-pm.com/crisp-dm-2/

S. Shanthi, & R. Geetha Ramani. (2012). *Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques*.

Sruthi ER. (2021). *Random Forest*. https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Sunil Ray. (2017). *Naive bayes*. https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

Thaduri, A., Polepally, V., & Vodithala, S. (2021). Traffic accident prediction based on CNN model. *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, 1590–1594. https://doi.org/10.1109/ICICCS51141.2021.9432224

V.A, O., & A.A, E. (2014). Traffic Accident Analysis Using Decision Trees and Neural Networks. *International Journal of Information Technology and Computer Science*, *6*(2), 22–28. https://doi.org/10.5815/ijitcs.2014.02.03

Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., & Wang, Z. (2019a). Traffic accident's severity prediction: A deep-learning approach-based CNN network. *IEEE Access*, *7*, 39897–39910. https://doi.org/10.1109/ACCESS.2019.2903319

Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., & Wang, Z. (2019b). Traffic accident's severity prediction: A deep-learning approach-based CNN network. *IEEE Access*, *7*, 39897–39910. https://doi.org/10.1109/ACCESS.2019.2903319

Annexure

This project report titled **Accident Severity Detection and Prediction Using Machine Learning Techniques** was scanned for similarity detection and the outcome is as given below.

Plagiarism score

Software Used: Turnitin

Date of Report Generation: 15th August 2022

Similarity Index in %: 16%

Total word count: 6288

PlagiarismReport

Annexure

Github Link to code

The following GitHub link contains the code base used for the project: Github link to Code files

REVA UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

DRIVE SAFE

ALERT TODAY…. ALIVE TOMORROW

Thank You!