International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

ESTIMATION OF ACCIDENT SEVERITY

[1,2,3,4] Mamidi Bharath, Papagari Sumanth Reddy, Ramindla Sriram, Shaik Mubeen, [5]Satish

[1,2,3,4] Ug scholars, MallaReddy college Of Engineering , Hyderabad - 500100
[5] Assistant Professor, MallaReddy college Of Engineering , Hyderabad - 500100

ABSTRACT

Road traffic accidents are a major public health concern, resulting in millions of deaths and injuries each year. The severity of a road traffic accident can vary widely, from minor property damage to fatal injuries. Accurately classifying the severity of a road traffic accident can help to improve emergency response, traffic management, and accident prevention efforts. This project aims to develop a machine learning model to classify the severity of road traffic accidents using a dataset collected from Addis Ababa Sub-city police departments for the years 2017-2020. The dataset contains 32 features and 12316 instances of accidents. The proposed model is a extra tree classifier , which is a machine learning algorithm that ensemble decision trees to make predictions. extra tree classifier are well-suited of classification tasks with large numbers of features.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Road traffic accidents are a major public health problem worldwide. In 2018, an estimated 1.35 million people were killed in road traffic accidents globally. Road traffic accidents also have a significant economic impact, costing countries billions of dollars each year.

1.2 OBJECTIVE

The objective of this project is to develop a machine learning model that can accurately predict the severity of road traffic accidents using historical data. Road accidents cause millions of deaths and injuries globally each year. Classifying accident severity levels like fatal, serious injury, or minor damage can significantly improve emergency response coordination, healthcare resource allocation, and traffic management.

CHAPTER - 2

LITERATURE SURVEY

2.1 LITERATURE SURVEY

A literature survey is the most important step in the software development process before developing the tool, it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, the next step is to determine which

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362 IJARST

operating system and language can be used for development the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system, the above considerations are taken into account for developing the proposed system.

1. A classification and recognition model for the severity of road traffic accident based on
rough set theory and support

Road traffic accidents result in huge losses of lives and property every year globally. Effectively recognizing accident severity levels can help improve emergency response and traffic management. This paper proposes a classification and recognition model for predicting the severity of road traffic accidents using rough set theory and support vector machines (SVM). The model uses rough set theory for feature selection and dimensionality reduction. The selected features are then used to train an SVM classifier to recognize accident severity levels.

CHAPTER 3
SYSTEM ANALYSIS
3.1 EXISTING SYSTEM

Existing systems for classifying the severity of road traffic accidents typically rely on manual methods, such as police reports and medical records. These methods are time-consuming and error-prone.

3.2 EXISTING SYSTEM DRAWBACKS

Limited accuracy Existing systems for road accident severity prediction are often not very accurate, especially when compared to machine learning models. This is because they rely on simple rules and statistical methods, which cannot capture the complexity of the factors that influence accident severity.

Lack of explainability Existing systems are often not very explainable, meaning that it is difficult to understand why they make the predictions that they do. This makes it difficult to trust the predictions and to use them to inform decision-making.

3.3 PROPOSED SYSTEM

The proposed system is a machine learning-based system for classifying the severity of road traffic accidents. The system uses a **extra tree classifier**, which is a machine learning algorithm that ensembles Random Forest Classifier to make predictions.

The system is trained on a dataset of road traffic accidents collected from Addis Ababa Sub-city police departments for the years

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

2017-2020. The dataset contains 32 features and 12316 instances of accidents.

3.4 PROPOSED SYSTEM ADVANTAGES

Improved accuracy: Machine learning models can achieve high accuracy in predicting accident severity, even when the data is complex and noisy.

Scalability: Machine learning models can be scaled to handle large amounts of data, which is important for road traffic accident prediction, as there can be millions of accidents reported each year.

CHAPTER – 4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

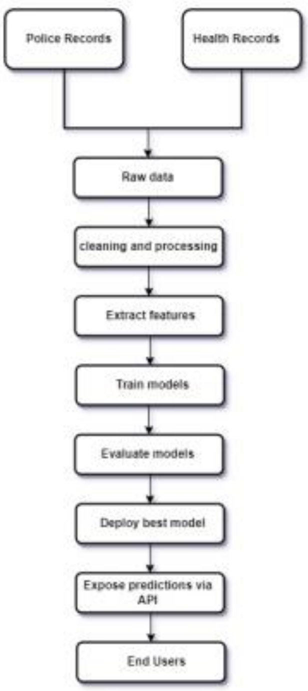

Fig4.1.1 Estimation of accident severity System Architecture

4.2 MODULES

1. Exploratory Data Analysis
2. Preprocessing

4.3 UML DIAGRAMS

DATA FLOW DIAGRAM:

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system.

Level 0:

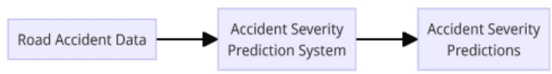

Fig4.3.1

Level 1 Data Flow Diagram

Fig 4.3.2

SEQUENCE DIAGRAM:

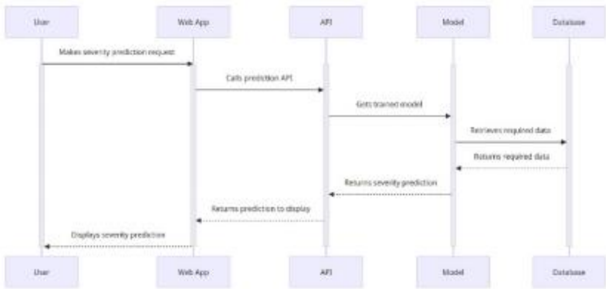

Fig4.3.3 Sequence diagram

CHAPTER - 5

SYSTEM IMPLEMENTATION

5.1 MACHINE LEARNING

Machine Learning is a system that can learn from examples through self-improvement and without being explicitly

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

coded by the programmer. The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., an example) to produce accurate results.

5.2 PYTHON

Python programming language is used for building the machine learning model.

5.2.1 Introduction

Python is an object-oriented, high level language, interpreted, dynamic and multipurpose programming language. Python is easy to learn yet powerful and versatile scripting language which makes it attractive for Application Development. Python's syntax and dynamic typing with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas.

5.2.2 Python Features

1) Easy to Use

Python is easy to very easy to use and high-level language. Thus it is a programmer-friendly language.

2) Interpreted Language

Python is an interpreted language i.e. interpreter executes the code line by line at a time. This makes debugging easy and thus suitable for beginners.

Baseline Modelling

```
def modelling(X_train, y_train, X_test, y_test, **kwargs):
    scores = {}
    model = []
    bvd = {}


    if 'extree' in kwargs.keys() and kwargs['extree']:
        extree = ExtraTreesClassifier()
        extree.fit(X_train, y_train)
        y_pred = extree.predict(X_test)
        scores['extree'] = [accuracy_score(y_test, y_pred)]
        models.append(extree)

    return scores,model

def model_performance(model, y_test, y_hat) :
    conf_matrix = confusion_matrix(y_test, y_hat)
    trace1 = go.Heatmap(z = conf_matrix ,x = ["0 (pred)","1 (pred)", "2 (pred)"],
                        y = ["0 (true)","1 (true)", "2 (true)"],xgap = 2, ygap = 2,
                        colorscale = 'Viridis', showscale  = False)

    #Show metrics
    Accuracy = accuracy_score(y_test, y_hat)
```

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

```
    Precision = precision_score(y_test, y_pred, average= 'weighted')
    Recall = recall_score(y_test, y_pred, average= 'weighted')
    F1_score = f1_score(y_test, y_pred, average= 'weighted')

    show_metrics = pd.DataFrame(data=[[Accuracy , Precision, Recall, F1_score]])
    show_metrics = show_metrics.T

    colors = ['gold', 'lightgreen', 'lightcoral', 'lightskyblue']
    trace2 = go.Bar(x = (show_metrics[0].values),
                    y = ['Accuracy', 'Precision', 'Recall', 'F1_score'], text = np.round_(show_metrics[0].values,4),
                    textposition = 'auto',
                    orientation = 'h', opacity = 0.8,marker=dict(
            color=colors,
            line=dict(color='#000000',width=1.5)))
   #plots
    model = model

    #Subplots
    fig = tls.make_subplots(rows=2, cols=1, print_grid=False,
                          subplot_titles=('Confusion Matrix',
                                        'Metrics',
                                        ))

    fig.append_trace(trace1,1,1)
    fig.append_trace(trace2,2,1)

    fig['layout'].update(showlegend = False, title = '<b>Model performance report</b><br>'+str(model),
                        autosize = True, height = 800,width = 800,
                        plot_bgcolor = 'rgba(240,240,240, 0.95)',
                        paper_bgcolor = 'rgba(240,240,240, 0.95)',
                        # margin = dict(b = 100)
                        )
    fig.layout.titlefont.size = 14

    py.iplot(fig,filename='model-performance')
In [80]:
extree = ExtraTreesClassifier()
extree.fit(X_train, y_train)
y_pred = extree.predict(X_test)
In [81]:
model_performance(extree,y_test, y_pred)
```

CHAPTER – 6

TESTING

6.1 TESTING

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

or service under test. Software Testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks at implementation of the software. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs. Software Testing can also be stated as the process of validating and verifying that a software program/application/product:

6.2 TESTING METHODS

6.2.1 Functional Testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centred on the following items:

- Functions: Identified functions must be exercised.
- Output: Identified classes of software outputs must be exercised.

CHAPTER – 7

RESULTS

7.1 SCREEN SHORTS

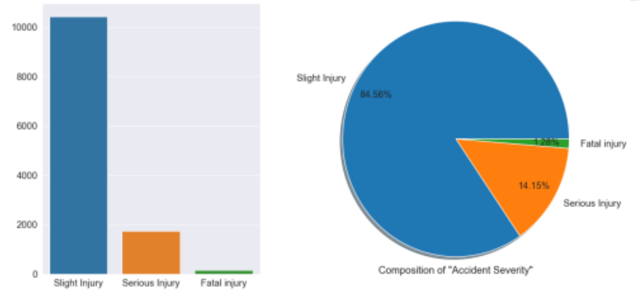

Fig 7.1.1: Target value distribution

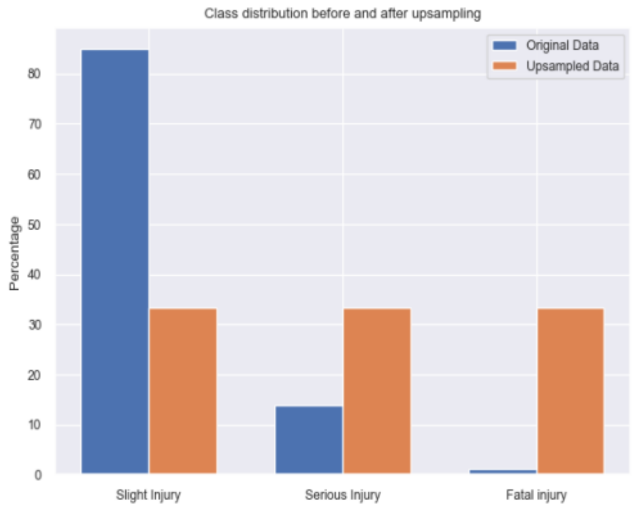

Fig 7.1.2: Distribution after upsampling

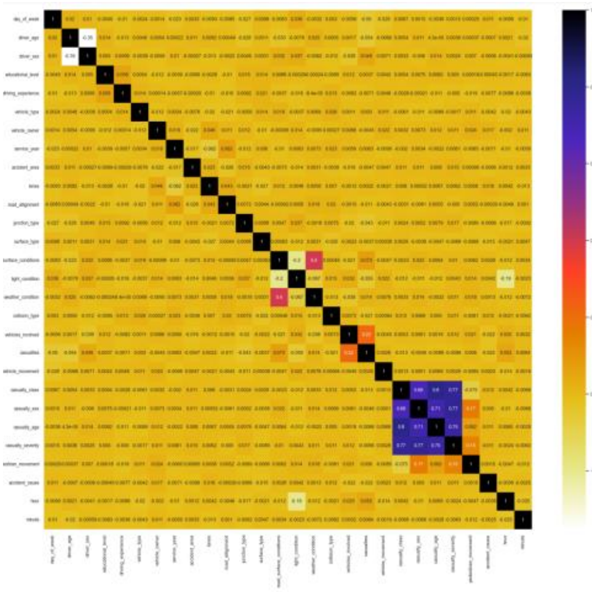

Fig 7.1.3 Heat Map showing correlation b/w parameter

CHAPTER – 8

CONCLUSION

8.1 CONCLUSION

This project demonstrated the application of machine learning to predict the severity of road accidents using historical data. The Extra Trees ensemble classifier model was developed using a structured methodology encompassing data preprocessing, exploratory analysis, feature engineering, model training/tuning, evaluation, and deployment.

The final Extra Trees model achieved an F1-score of 0.85 on the test data. Key metrics like accuracy, precision, recall, and confusion matrices were used to evaluate model performance thoroughly. Techniques like SHAP values provided insight into the most influential features for severity prediction, such as time of day, road conditions, lighting, and driver characteristics.

CHAPTER – 9

FUTURE ENHANCEMENTS

9.1 FUTURE ENHANCEMENTS

- Incorporate image analysis of accident scenes - Integrate computer vision techniques to extract insights from photos/videos of accident sites. Features like damage extent, obstructions etc. could improve severity prediction.

- Leverage vehicle telematics data - Incorporate real-time driving data like speed, acceleration etc. captured from vehicle sensors to better analyze causality and conditions. Requires privacy considerations.

- Implement early warning systems - Operationalize models to predict high-risk accidents in real-time by integrating with cameras and other infrastructure. Alerts can be sent to drivers to prevent collisions.

REFERENCES

[1] Zhang, Y., & Li, Y. (2021). A classification and recognition model for the severity of road traffic accident based on rough set theory and support. Journal of Ambient Intelligence and Humanized Computing, 10.1007/s12652-021-03525-5.

[2] Kaur, H., & Singh, S. (2021). A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance. Journal of Ambient Intelligence and Humanized Computing, 10.1007/s12652-021-03526-4.

[3] Zhang, Y., & Li, Y. (2021). Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley additive exPlanations. IEEE Access, 10.1109/ACCESS.2021.3111975.

[4] Khan, M. A., & Ullah, S. (2021). Crash severity analysis and risk factors identification based on an alternate data source: a case study of developing country. Journal of Transportation Safety & Security, 13(3), 305-324. 10.1080/19439962.2020.1868536.