



 Edna722 / ML-Projct Public GPL-3.0 license 0 stars  0 forks Star Watch[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#) main ▾

...

Edna722 Merge branch 'main' of <https://github.com/Edna722/ML-Projct> ...1 hour ago [View code](#) README.md 

ML-Projct

Project Details:

Data Source: The data used for this project is obtained from the Chicago Police Department Crash Dataset. It includes various parameters related to traffic crashes such as date, street name, weather condition, posted speed limit, traffic way type, and types of road defects.

Goals:

The goal of this project is to create a model that predicts and classifies different types of traffic crashes, with a specific focus on predicting "Rear End" crashes. The dataset parameters are utilized to distinguish between various types of crashes, including "Rear End" crashes.

Problems:

During the project, the following challenges were encountered:

- Applying label encoding and one-hot encoding techniques and merging dataframes to obtain a comprehensive dataset with all the required information.
- Dealing with the loss of data after merging or adding columns.
- Handling a significant number of "Unknown" values and NaNs in the dataset.



Solutions:

- To address these challenges, the following approaches were taken:
- Conducting effective online searches for assistance and guidance on specific issues.
- Experimenting with the code and making necessary modifications to ensure proper functionality and data handling.
- Recommendations for further developments:
- Based on the project findings, the following recommendations are suggested for further development:
- Incorporate additional datasets related to vehicles and drivers/passengers and combine them with the existing Traffic Crash dataset for a more comprehensive analysis.
- Explore the possibilities of binning and clustering the data to identify crash locations and patterns.
- Continuously work on improving the predictive model's accuracy and performance.

Project Intro/Objective:

- The main objective of this project is to determine whether a crash is a "Rear End" crash or not. To achieve this, DecisionTreeClassifier and RandomForestClassifier models are used to classify the crashes based on various parameters, such as traffic control devices, weather conditions, lighting conditions, roadway conditions, and contributory causes.

Project Findings:

- Based on the analysis conducted, the most important feature in classifying the crashes was found to be the "PRIM_CONTRIBUTORY_CAUSE" parameter. The specific causes, such as "Following too closely," "Failing to yield right-of-way," and "Improper lane usage," played significant roles in determining the crash type.

Descriptive statistics

- Data cleaning and preprocessing
- Data exploration and visualization
- Label encoding and one-hot encoding
- Feature engineering
- Machine learning algorithms (DecisionTreeClassifier and RandomForestClassifier)
- Model evaluation using accuracy_score and roc_auc_score metrics
- GridSearchCV for hyperparameter tuning
- Visualization of decision tree using export_graphviz

Python programming language

- Pandas and NumPy for data manipulation and analysis
- Jupyter Notebook for code development and documentation
- Seaborn and Matplotlib for data visualization
- Scikit-learn for machine learning algorithms and evaluation
- Needs of this project:
- The project addresses the following needs:

Exploration and analysis of car crash data

- Data processing and cleaning
- Statistical modeling and prediction
- Report writing and documentation
- Identification of parameters leading to crashes
- Understanding the time and date patterns of crashes

Modelling

- The models used include RandomClassifier, Grid Search amongst others so as to give correct predictions

Project Findings:

- The most important feature that played a big role in classifying was 'PRIM_CONTRIBUTORY_CAUSE'.
 - 'UNABLE TO DETERMINE', 'FAILING TO YIELD RIGHT-OF-WAY',
- 'FOLLOWING TOO CLOSELY', 'NOT APPLICABLE',
- 'IMPROPER OVERTAKING/PASSING', 'IMPROPER BACKING',
- 'FAILING TO REDUCE SPEED TO AVOID CRASH', 'IMPROPER LANE USAGE',
- 'IMPROPER TURNING/NO SIGNAL', 'DRIVING SKILLS/KNOWLEDGE/EXPERIENCE',
- 'WEATHER', 'DISREGARDING TRAFFIC SIGNALS', etc.

Suggestions for future advancements

- Additionally, considering binning and clustering techniques can be beneficial for analyzing crash locations and identifying patterns or clusters of crashes in specific areas. This can offer insights into geographical aspects related to traffic accidents.
- Furthermore, there is room for improvement in the model itself. It is advisable to explore and implement enhancements to enhance the model's performance and predictive accuracy. This may involve experimenting with different algorithms, feature engineering techniques, hyperparameter tuning, or exploring ensemble methods to achieve better results.



Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%

