Victor-Mufasa / **Phase-4-Project**

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Watch 0

Deep Machine Leaning Project

0 stars 0 forks 0 watching Branches Activity Tags

Public repository

7 Branches 0 Tags Go to file t Go to file Add file Code

justshron Remove file from repo 8d9ac4f · 8 minutes ago

data	Added the three datasets	last week
notebooks	initial commit	last week
Chicago Crashes ppt.pdf	Renamed file from 'Chicago Crashes...	13 minutes ago
README.md	add relevant links	6 hours ago
index.ipynb	improved 2nd model accuracy	6 hours ago

README

Road Traffic Accident Analysis: Predictive Modeling and Safety Insights

Project Overview

This project analyzes road traffic crash data to identify key contributors to accidents and predict accident severity using machine learning. By combining exploratory data analysis with XGBoost classification models, the project provides actionable insights for improving road safety and reducing accident-related injuries and fatalities.

Objectives

- **Explore** road traffic crash patterns and trends through comprehensive data analysis
- **Develop** a predictive machine learning model to classify primary contributory causes of accidents
- **Evaluate** model performance using accuracy, precision, recall, and F1-score metrics

- **Provide** data-driven recommendations for road safety improvements

Dataset

The analysis utilizes three interconnected datasets:

- **Traffic_Crashes_-_Crashes.csv.gz**: 54,959 crash records with 48 features including weather, lighting, road conditions, and injury severity
- **Traffic_Crashes_-_People.csv.gz**: Information about individuals involved in crashes
- **Traffic_Crashes_-_Vehicles.csv.gz**: Vehicle-specific data for crashes

Key Features

Exploratory Data Analysis

- Time-based analysis (hourly, daily, monthly crash patterns)
- Geographic analysis of high-risk streets and locations
- Weather and lighting condition impact assessment
- Speed limit and roadway surface condition analysis
- Statistical hypothesis testing (ANOVA) for environmental factors

Feature Engineering

- Speed limit discretization into categorical bins
- Severity classification based on injury levels
- Aggregation of vehicle and people data
- Rare category grouping for improved model performance

Machine Learning Models

- **Random Forest Classifier**: Baseline model with balanced class weights
- **XGBoost Classifier**: Advanced gradient boosting with sample weight optimization
- Multi-class classification of primary contributory causes
- Class imbalance handling through computed sample weights

Key Findings

Accident Patterns

- **Peak Times**: Most crashes occur between 3 PM - 6 PM during rush hour
- **High-Risk Days**: Friday and Tuesday show highest crash frequencies (8,200+ crashes)
- **Seasonal Trends**: Summer months (July-August) have peak crash rates; December has the lowest
- **Top Streets**: Western Avenue (1,579 crashes), Cicero Avenue, and Pulaski Road are highest-risk locations

Contributing Factors

- **Primary Causes**:
    - Failure to yield right-of-way
    - Following too closely
    - Improper lane usage
- **Environmental Impact**: Wet roadway surfaces show 2.3% severe injury rate (highest among conditions)
- **Speed Correlation**: Speed limits of 60-65 km/h show 20% probability of severe injuries
- **Crash Types**: 71.5% result in property damage only; 28.5% involve injury/fatality

Statistical Significance

- Weather conditions significantly affect injury rates ($p < 0.01$)
- Lighting conditions significantly impact accident severity ($p < 0.01$)

Model Performance

XGBoost Classifier (Final Model)

```
Overall Accuracy: 36%

Class-wise Performance:
- VIOLATION:      Precision: 0.36, Recall: 0.70, F1: 0.47
- UNKNOWN:        Precision: 0.69, Recall: 0.32, F1: 0.44
- DRIVER_ERROR:   Precision: 0.42, Recall: 0.27, F1: 0.33
- ENVIRONMENT:    Precision: 0.15, Recall: 0.87, F1: 0.25
```

The model demonstrates strong recall for violations and environmental factors, indicating effective identification of these accident causes despite class imbalance challenges.

Technologies Used

- **Python 3.x**
- **Data Analysis**: pandas, numpy
- **Visualization**: matplotlib, seaborn
- **Statistical Testing**: scipy
- **Machine Learning**:
    - scikit-learn (preprocessing, metrics, Random Forest)
    - XGBoost (gradient boosting classifier)
    - imbalanced-learn (SMOTE, class balancing)

Project Structure

```
├── data/
│   ├── Traffic_Crashes_-_Crashes.csv.gz
│   ├── Traffic_Crashes_-_People.csv.gz
│   └── Traffic_Crashes_-_Vehicles.csv.gz
```

```
├── index.ipynb
├── README.md
```

Installation & Usage

```
# Clone the repository
git clone <git@github.com:Victor-Mufasa/Phase-4-Project.git>

# Install required packages
pip install pandas numpy matplotlib seaborn scipy scikit-learn xgboost imbalanced-learn

# Run the Jupyter notebook
jupyter notebook index.ipynb
```

Recommendations

Based on the analysis, the following interventions are recommended:

1. Targeted Traffic Enforcement

- Deploy automated enforcement at high-violation intersections
- Increase monitoring during peak hours (3-6 PM)
- Focus on Western Avenue, Cicero Avenue, and other high-risk streets

2. Driver Education Programs

- Mandatory refresher courses on yielding, following distance, and lane usage
- Public awareness campaigns targeting common driver errors
- Special training for high-risk behaviors

3. Environmental Interventions

- Variable speed limits adjusted for weather conditions
- Enhanced road maintenance for wet/icy conditions
- Improved drainage systems to reduce wet surface accidents

4. Infrastructure Improvements

- Upgrade lighting at accident-prone locations
- Regular traffic control device maintenance
- Enhanced signage at high-risk intersections

Future Work

- Incorporate real-time weather data for dynamic risk assessment

- Develop deep learning models for improved prediction accuracy
- Create interactive dashboard for traffic safety monitoring
- Implement geospatial clustering for accident hotspot identification
- Explore time-series forecasting for proactive intervention planning

Acknowledgments

Data sourced from [City/Municipality] Traffic Safety Division

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Contributors 6

Languages

● **Jupyter Notebook** 100.0%