

oblomovite /
chicago_car_crash_ml

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

This project is for the phase three flatiron school project

0 stars

0 forks

1 watching

Branches

Activity

Tags

Public repository

1 Branch

0 Tags

Go to file

Go to file

+

Add file

Code

oblomovite +finalize

6f8beed · now

| | | |
|------------------|--|-------------|
| .DS_Store | add business understanding, busin... | 2 weeks ago |
| .gitignore | use pipeline, refactor preprocess l... | last week |
| README.md | +finalize | now |
| index.ipynb | +finalize | now |
| requirements.txt | create README and requirements | last week |

README

Crash Severity Analysis: A Data-Driven Approach

This project applies machine learning to analyze crash severity factors using logistic regression. The goal is to identify key risk factors contributing to severe crashes and provide business recommendations for improving road safety.

Usage

1. Install dependencies and place all CSV data in the /data folder.
2. Install the required packages using the following command:

```
pip install -r requirements.txt
```
3. Run the supplied notebook to preprocess data, train models, and view performance metrics and checkpoints.

Business Understanding

- Determine what factors contribute to severe crashes and how they can be mitigated.
- Key stakeholders include Traffic Safety Departments, Urban Planners, and Traffic Authorities.
- Recommendations may improve traffic safety measures, vehicle regulations, and driver behavior enforcement.

Data Understanding

Source

- Sourced from Chicago's open data portal, merging the following records using CRASH_RECORD_ID:
 - [Crash](#)
 - [Vehicle](#)
 - [People](#)

Features

- Imbalanced classes, with fewer fatal crashes compared to non-fatal ones.
- Multiple categorical variables requiring encoding for machine learning.
- Missing and unknown values in some safety equipment and environmental factors.
- Robust and diverse feature set with 650,000+ records and 50+ features.

Data Preparation

- Feature engineering:
 - Binned continuous variables to reduce dimensionality.
 - Reducing the number of unique values in categorical features.
 - One-hot encoding categorical variables.
 - Standardized numerical variables.
- Missing data handling:
 - Used mode imputation for categorical variables.
 - Dropped irrelevant or highly missing columns.
- Addressing class imbalance:
 - Used random under-sampling (RUS) to balance fatal and non-fatal crashes.

Modeling

- Applied Logistic Regression and Decision Tree models to predict crash severity.
- Extracted regression coefficients to determine the most influential factors in crash severity.

Evaluation

- F1 score, precision, recall, and accuracy metrics assess model performance.
- ROC-AUC, and cross-validation evaluate model performance.
- Confusion matrices and classification reports provide detailed performance metrics.
- Cross-validation

Findings

- **Driver's Physical Condition Significantly Affects Crash Severity**

- PRIM_CONTRIBUTORY_CAUSE_PHYSICAL CONDITION OF DRIVER (**Coefficient: 3.04**)

Actionable Steps:

- Implement roadside impairment tests using AI-powered assessment tools.
- Develop awareness campaigns on the risks of driving with medical impairments.

- **Speeding is a Major Predictor of Severe Crashes**

- PRIM_CONTRIBUTORY_CAUSE_EXCEEDING AUTHORIZED SPEED LIMIT (**Coefficient: 2.79**)

Actionable Steps:

- Expand speed camera enforcement in high-risk areas.
- Implement dynamic speed limits based on traffic and weather conditions.
- Increase public awareness campaigns about the dangers of excessive speed.

- **Pedestrian, Cyclist, Animal Crashes Have High Severity**

- FIRST_CRASH_TYPE_PEDESTRIAN (**Coefficient: 2.55**)
- FIRST_CRASH_TYPE_PEDALCYCLIST (**Coefficient: 1.39**)
- FIRST_CRASH_TYPE_ANIMAL (**Coefficient: 2.428**)

Actionable Steps:

- Improve car traffic isolation from bike lanes and pedestrian crossings. - Implement wildlife detection systems to alert drivers of animal crossings. - Increase visibility and signage at pedestrian and cyclist crossings. - Implement wildlife crossing deterrents (eg. fencing) in high-risk areas.

- **Dangerous Roadway Conditions Increase Crash Severity**

- ROAD_DEFECT_WORN SURFACE (**Coefficient: 0.97**)
- ROAD_DEFECT_RUT, HOLES (**Coefficient: 0.51**)
- ROADWAY_SURFACE_COND_SNOW OR SLUSH (**Coefficient: 0.61**)

Actionable Steps:

- Prioritize road maintenance funding for fixing potholes and worn surfaces.

- Implement real-time road hazard alerts using IoT and traffic apps.
- Expand pre-winter road treatment programs to reduce ice-related crashes.

- **High-Risk Intersection Designs Need Attention**

- TRAFFICWAY_TYPE_Y-INTERSECTION (**Coefficient:** 1.28)
- TRAFFICWAY_TYPE_T-INTERSECTION (**Coefficient:** 1.50)
- TRAFFICWAY_TYPE_UNKNOWN INTERSECTION TYPE (**Coefficient:** 1.93)
- TRAFFICWAY_TYPE_L-INTERSECTION (**Coefficient:** 1.34)

Actionable Steps:

- Improve signage and lane markings for better driver guidance.
- Deploy collision-prevention warning systems at intersections.
- Implement red-light cameras at high-risk intersections. - Implement traffic slowing measures at non-standard or high-risk intersections.

- **Distracted and Reckless Driving Remains a Major Threat**

- PRIM_CONTRIBUTORY_CAUSE_DISTRACTION - FROM OUTSIDE VEHICLE (**Coefficient:** 1.26)
- PRIM_CONTRIBUTORY_CAUSE_OPERATING VEHICLE IN ERRATIC, RECKLESS, CARELESS, NEGLIGENT OR AGGRESSIVE MANNER (**Coefficient:** 1.66)
- PRIM_CONTRIBUTORY_CAUSE_CELL PHONE USE OTHER THAN TEXTING (**Coefficient:** 0.42)

Actionable Steps:

- Implement stricter penalties for distracted and reckless driving.
- Launch public awareness campaigns on the dangers of distracted driving.
- Install CCTV cameras at high-risk intersections to monitor driver behavior and issue fines.

Project Structure

```
|— data/           # project data
|— checkpoint/    # data checkpoints
|— index.ipynb    # main notebook
|— README.md      # project overview
```



Next Steps

- Fine-tune threshold techniques to balance classes effectively.



No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%