Accident Severity Prediction with Machine Learning

Moataz Bellah Ben Khedher

Introduction

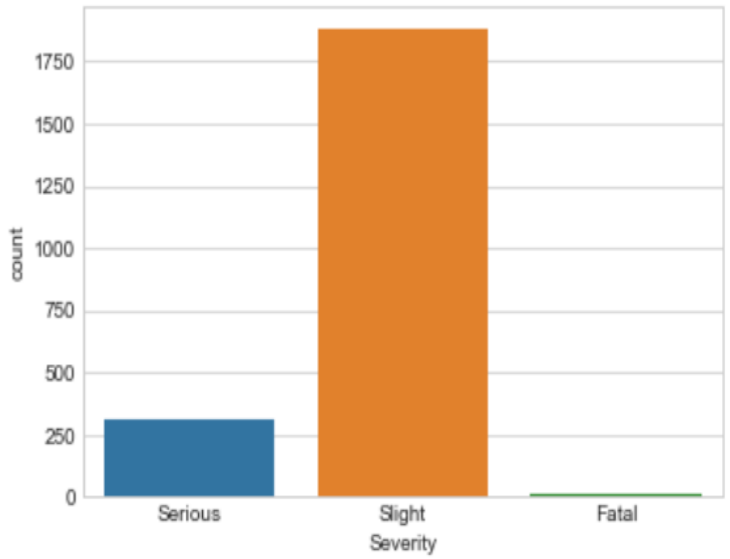
- About 1.25 million deaths per year
- Accidents are the Leading cause of death among young people aged 15 to 29
- Almost half of those killed on the roads are “vulnerable users” (pedestrians, cyclists and motorcyclists).
- Without sustained action, road accidents are projected to become the seventh leading cause of death by 2030.
- Accurate prediction of accident severity can be helpful to provide proactive solutions and test the readiness of road practitioners and local governments.
- In this project, we develop machine learning models to predict severity of road accidents in UK based on data collected in 2017.

Data

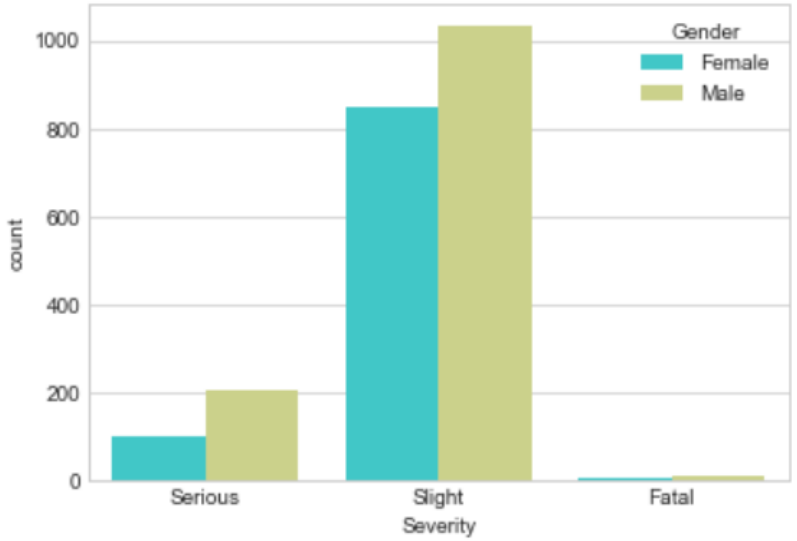
- Data acquired from UK government open data portal website
- Dataset was downloaded as CSV file
- 2203 records of traffic accidents that occurred in Leeds, UK in 2017

Data Exploration

Data count

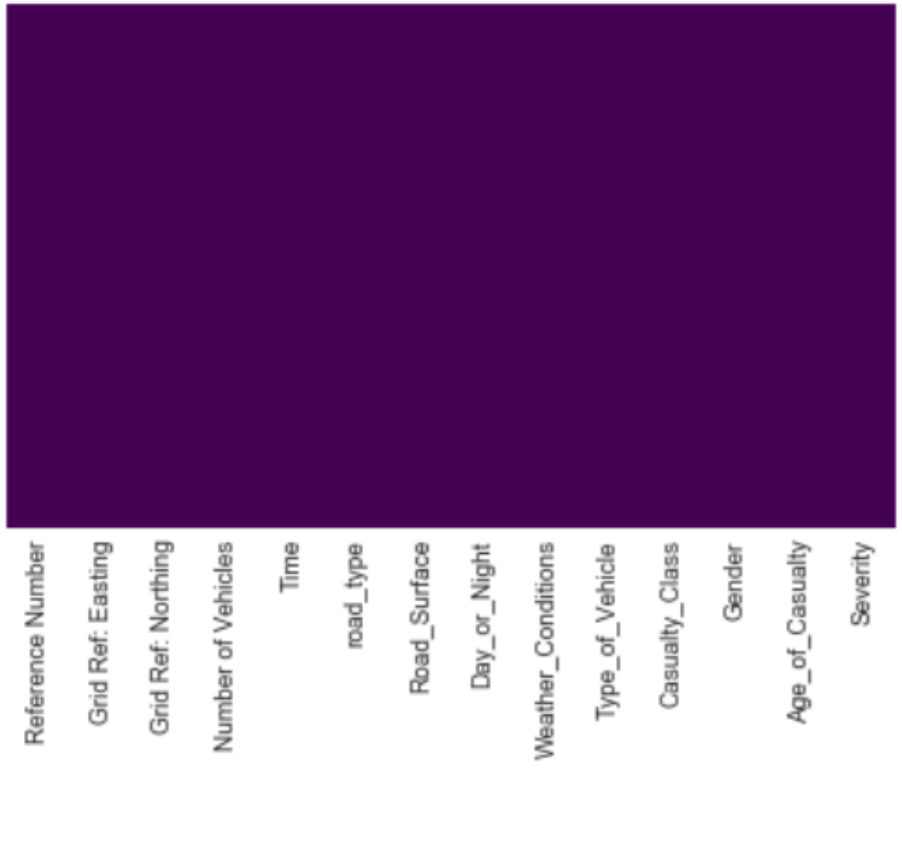


Severity by Gender



Data preprocessing

- Check for Missing Data



Data preprocessing

- Removing irrelevant attributes

	Time	road_type	Road_Surface	Day_or_Night	Weather_Conditions	Casualty_Class	Gender	Age_of_Casualty	Severity
0	815	A	Dry	Day	Other	Pedestrian	Female	61	Serious
1	1330	A	Dry	Day	Fine without high winds	Driver or rider	Male	36	Slight
2	805	A	Wet/Damp	Day	Fine without high winds	Driver or rider	Male	32	Slight
3	805	A	Wet/Damp	Day	Fine without high winds	Driver or rider	Male	30	Slight
4	1705	U	Wet/Damp	Night	Raining without high winds	Vehicle or pillion passenger	Female	26	Slight

Data preprocessing

- Feature Engineering: Convert time to categorical variable : Peak and Offpeak

	Time	road_type	Road_Surface	Day_or_Night	Weather_Conditions	Casualty_Class	Gender	Age_of_Casualty	Severity
0	peak	A	Dry	Day	Other	Pedestrian	Female	61	Serious
1	offpeak	A	Dry	Day	Fine without high winds	Driver or rider	Male	36	Slight
2	peak	A	Wet/Damp	Day	Fine without high winds	Driver or rider	Male	32	Slight
3	peak	A	Wet/Damp	Day	Fine without high winds	Driver or rider	Male	30	Slight
4	peak	U	Wet/Damp	Night	Raining without high winds	Vehicle or pillion passenger	Female	26	Slight

Model Building

Build 6 machine learning classifiers:

- Logistic Regression (LR)
- LinearDiscriminantAnalysis (LDA)
- K-Nearest Neighbor (Knn)
- Decision Tree (CART)
- Gaussian Naive Bayes (NB)
- Support Vector Machine (SVM)

Model Building

- Data were splitted into 80% training data and 20% validation data

Model	Accuracy
LR	0.8524
LDA	0.8411
KNN	0.8382
CART	0.7434
NB	0.0590
SVM	0.8536

- SVM slightly outperforms LDA, so we choose it for further validation

Model Building

	Precision	Recall	F1-Score	Support
Fatal	0.00	0.00	0.00	2
Serious	0.00	0.00	0.00	64
Slight	0.85	1.00	0.92	375
Avg/total	0.72	0.85	0.78	441

Conclusion

- In this project, we propose a machine learning model to predict the severity of traffic accidents in the city of Leeds, UK.
- A comparative approach was followed to identify the best performing model among 6 models.
- We found that SVM model performs the best but not the optimal
- Critical imbalance between the 3 severity classes
- Further improvements are required to optimize the model performance.