

Coursera Capstone project

ACCIDENT SEVERITY PREDICTION IN UK



Background

- Car accidents are one of the biggest cause of injuries and deaths all over the world.
- They impact economy with increased
 - commuting times
 - delivery times
 - Costs
 - pollution

The problem

❖ *Building a model to predict car accident severity can help commuters, professional drivers or logistic planners to reduce the personal and/or business impact of car accidents*

Stakeholders

- Individuals
 - work commuters
 - Taxi drivers
 - truck drivers
 - bus drivers §
- Businesses § logistic companies
 - public/private passengers bus companies
 - Taxi companies
- Government agencies
 - urban/suburban mobility managers

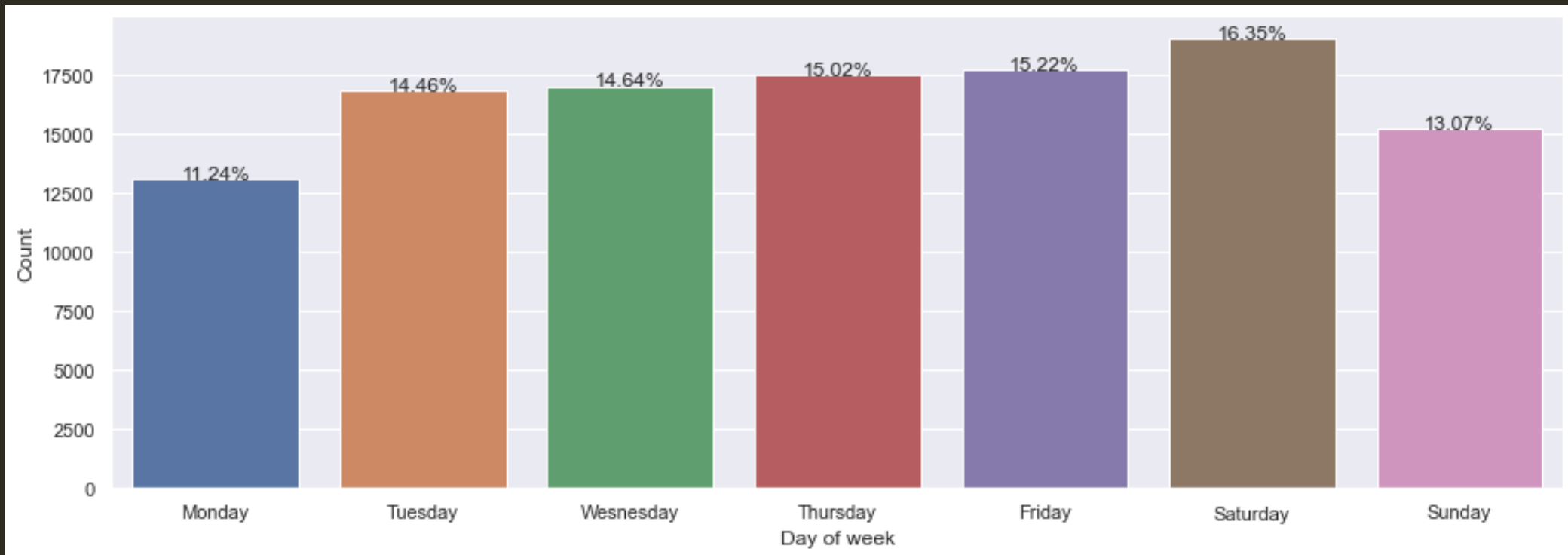
DATA SOURCES

- *Accident data was collected from UK Department of Transport downloaded in csv format*
- *Around 200k events*

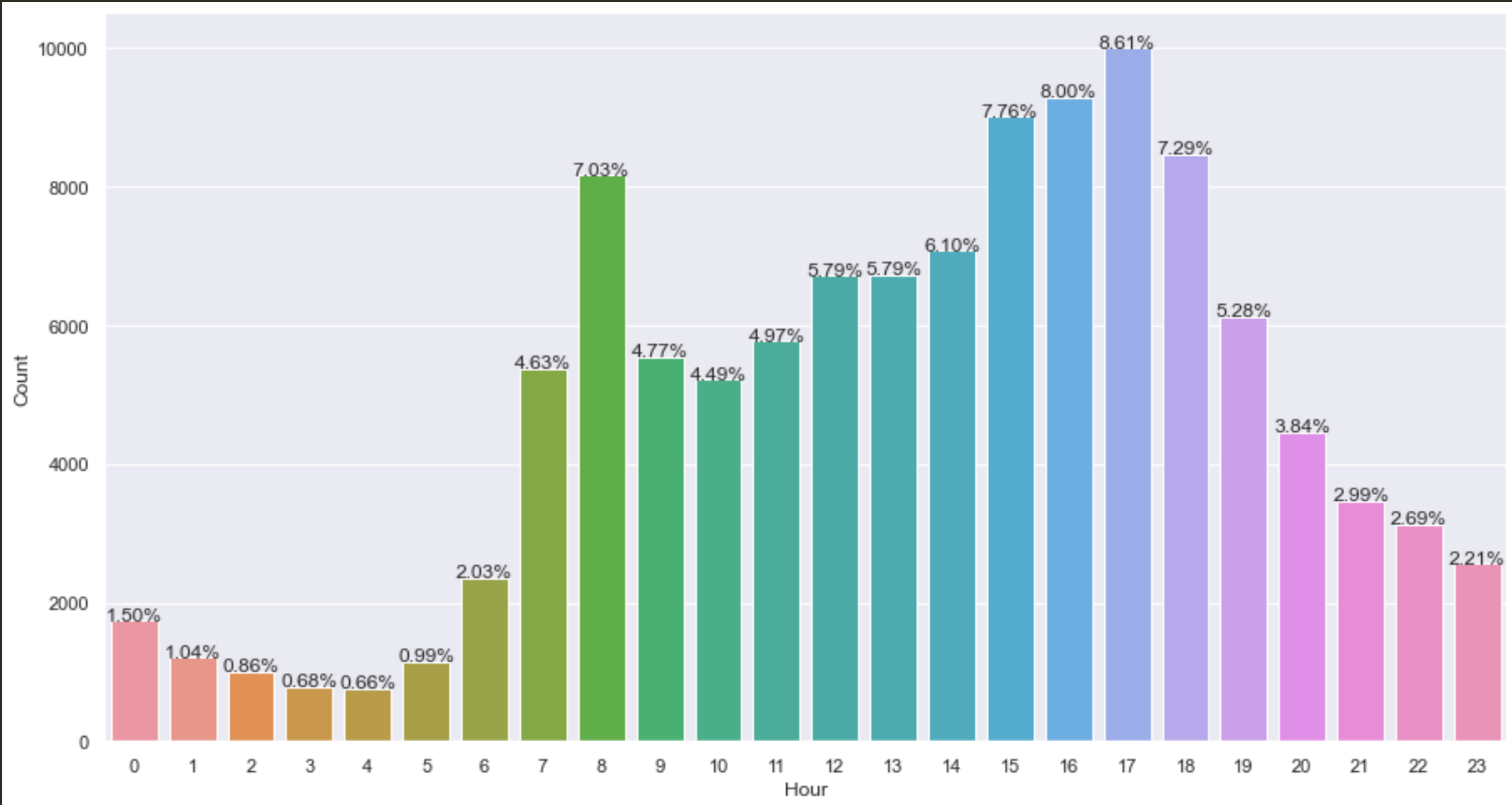
Feature Selection

- Day of Week
- Light Conditions
- Weather Conditions
- Road Surface Conditions
- Carriage Inwards

Relationship between Day of week and number of Accidents



Relationship Between Hours and Accidents



Models

- k-nearest Neighbours algorithm
- Decision Tree
- Support Vector Machine
- Logistic Regression

Evaluation

- Confronting the metrics of the different models, we look for the highest F1 score, the largest Jaccard score and the smallest log loss. Knn model is selected

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.251751	0.402115	NA
SVM	0.250214	0.399372	NA
DecisionTree	0.244702	0.391529	NA
Logistic Regression	0.210959	0.333193	1.08484

Conclusion

- *The model can non be used to predict the severity of an accident*
- *Accidents seem to overly take place in the morning at 8 and evenings, possibly when people, Normally goes to office and come back from office*
- *From observable features, the stakeholders can make an informed decision*