# Prediction of severity of road accidents using multiclass classification
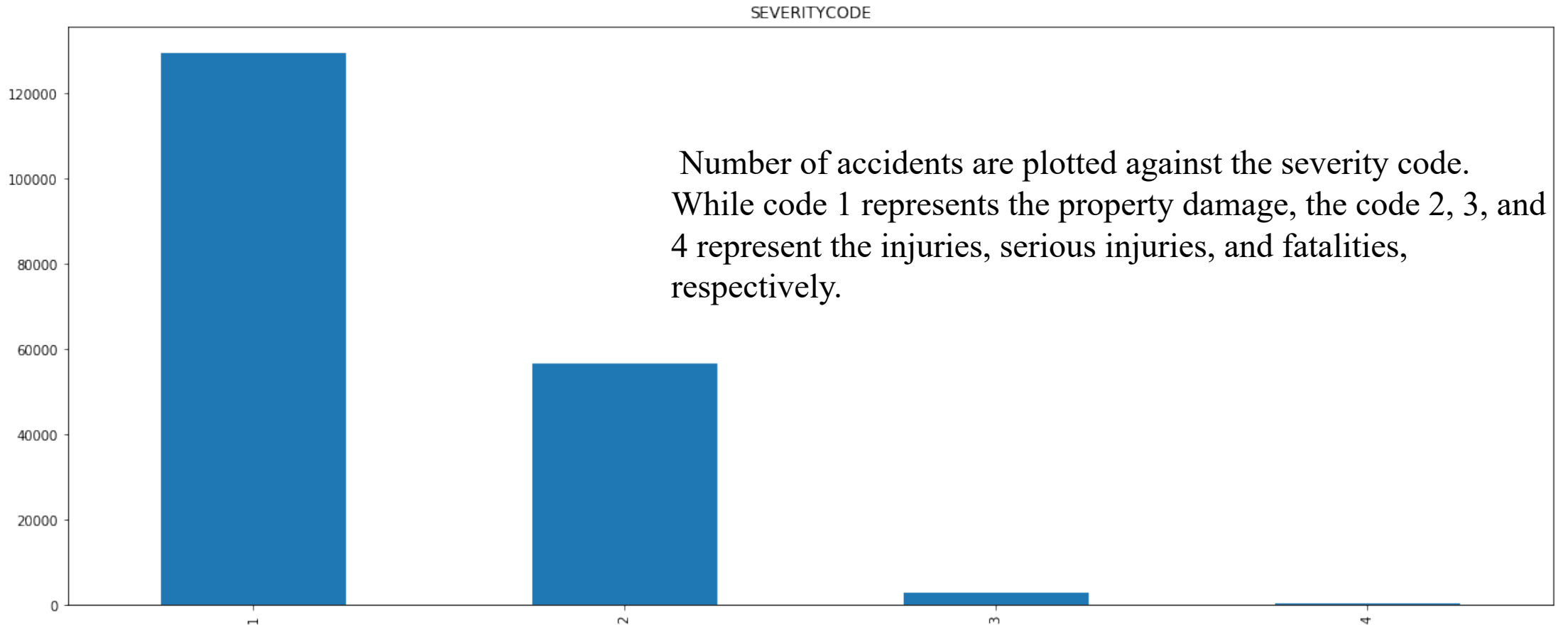
Lakshika Girihagama

# Introduction

- How to predict accurately severity of road collisions ahead remains one of the key questions for many.

- Road accidents severity includes property damages to injuries and ultimately unfortunate fatalities.

- Severity of road collisions can be influenced by various factors including personal behaviors (e.g. drugs under influence-DUI, speeding), environmental conditions (e.g. rain), and technological factors (e.g. car breakdown).

- Accurate prediction of severity of road accidents and their influencing factors can help people and authorities to adapt precautionary safety measures so that avoiding the risk may possible.

In this presentation, I address the question of how to accurately predict severity of road accidents using multiclass classification machine learning models.

# Data

- The data set contains the all types of road collisions recorded in the city of Seattle.

- This data set includes 221266 records of all types of collisions, their severity, various sub-categories of conditions (37) involving the accident.

- The time span for this data set ranges from 2004 to Present.

- The data record is updated weekly.

# Severity of accidents

SEVERITYCODE



Number of accidents are plotted against the severity code. While code 1 represents the property damage, the code 2, 3, and 4 represent the injuries, serious injuries, and fatalities, respectively.
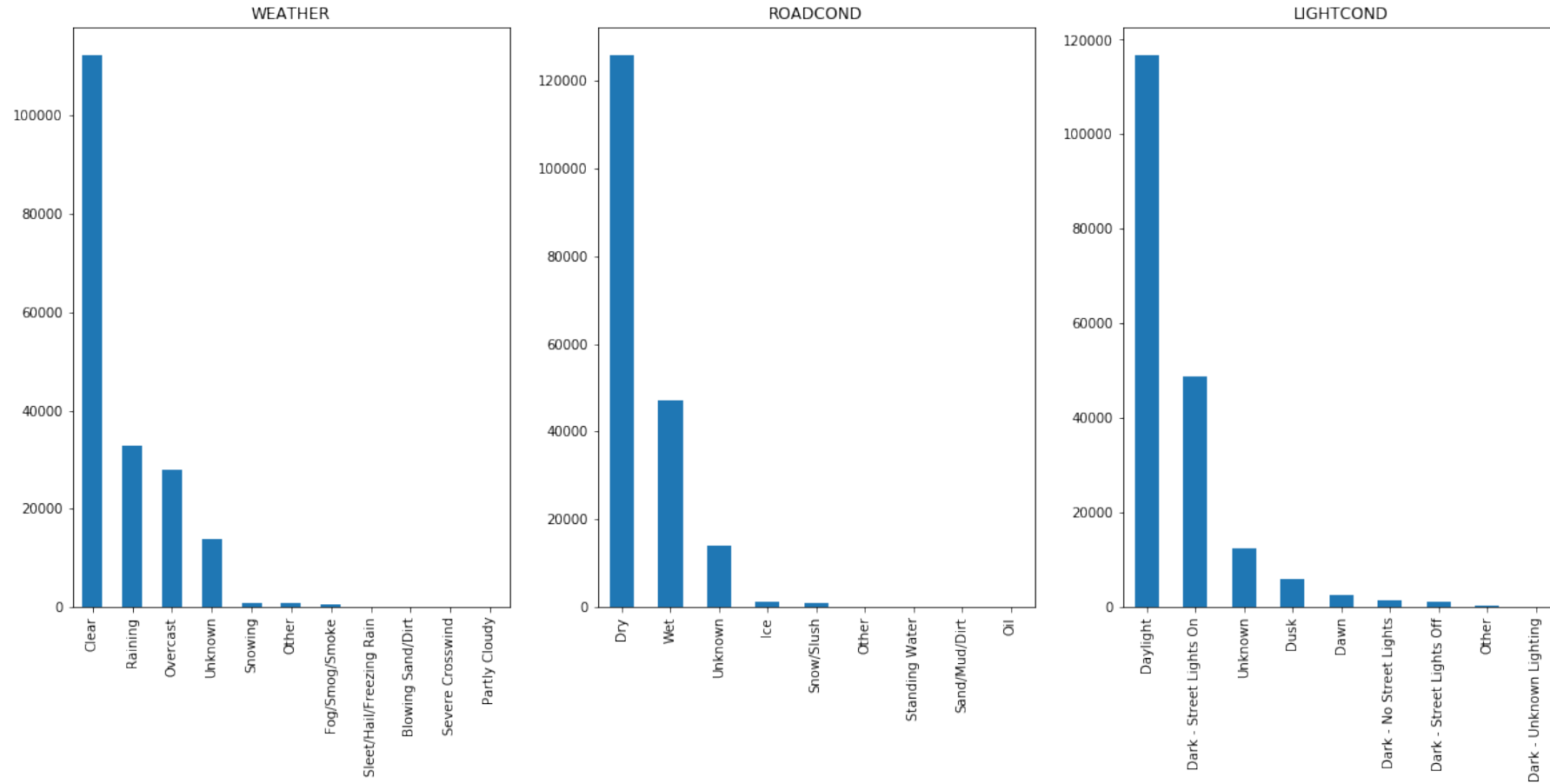
Distribution of number of accidents with respect to the day of week. 0: Monday, 1: Tuesday, 2: Wednesday, 3: Thursday, 4: Friday, 5: Saturday, and 6: Sunday.
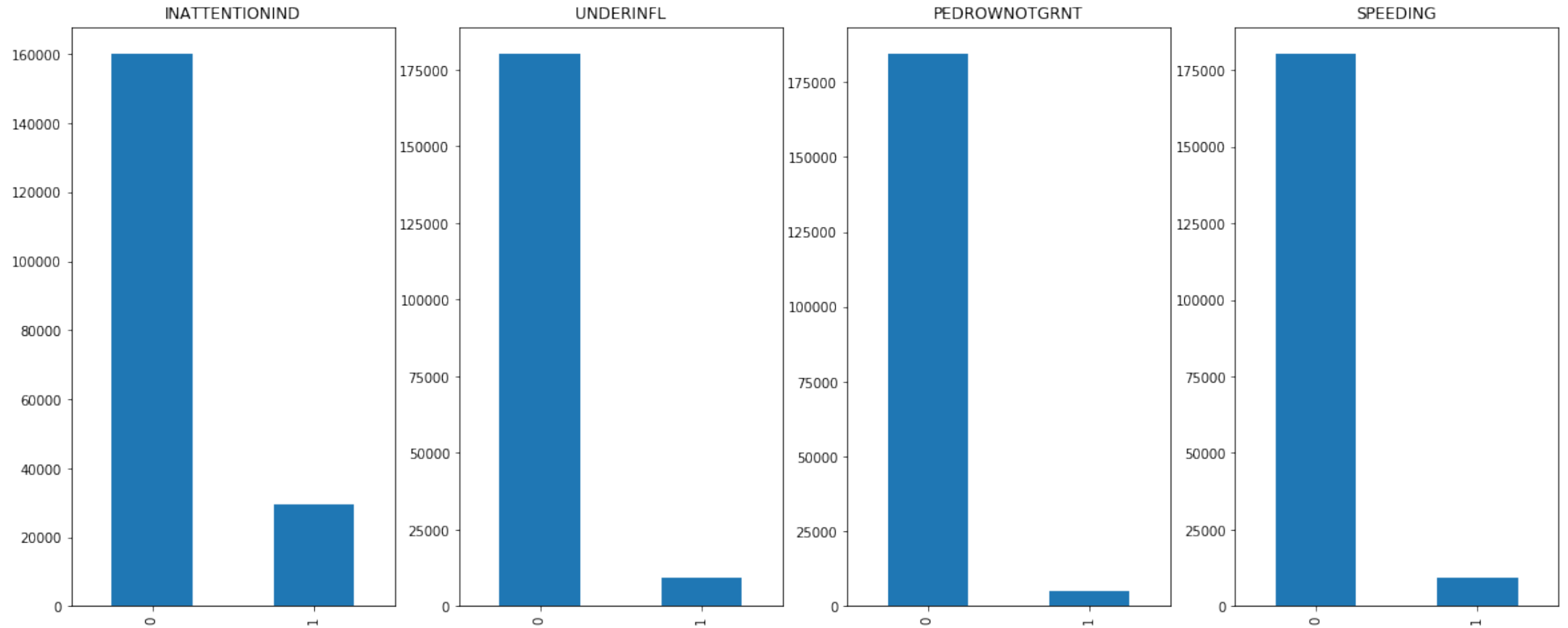


Day of week
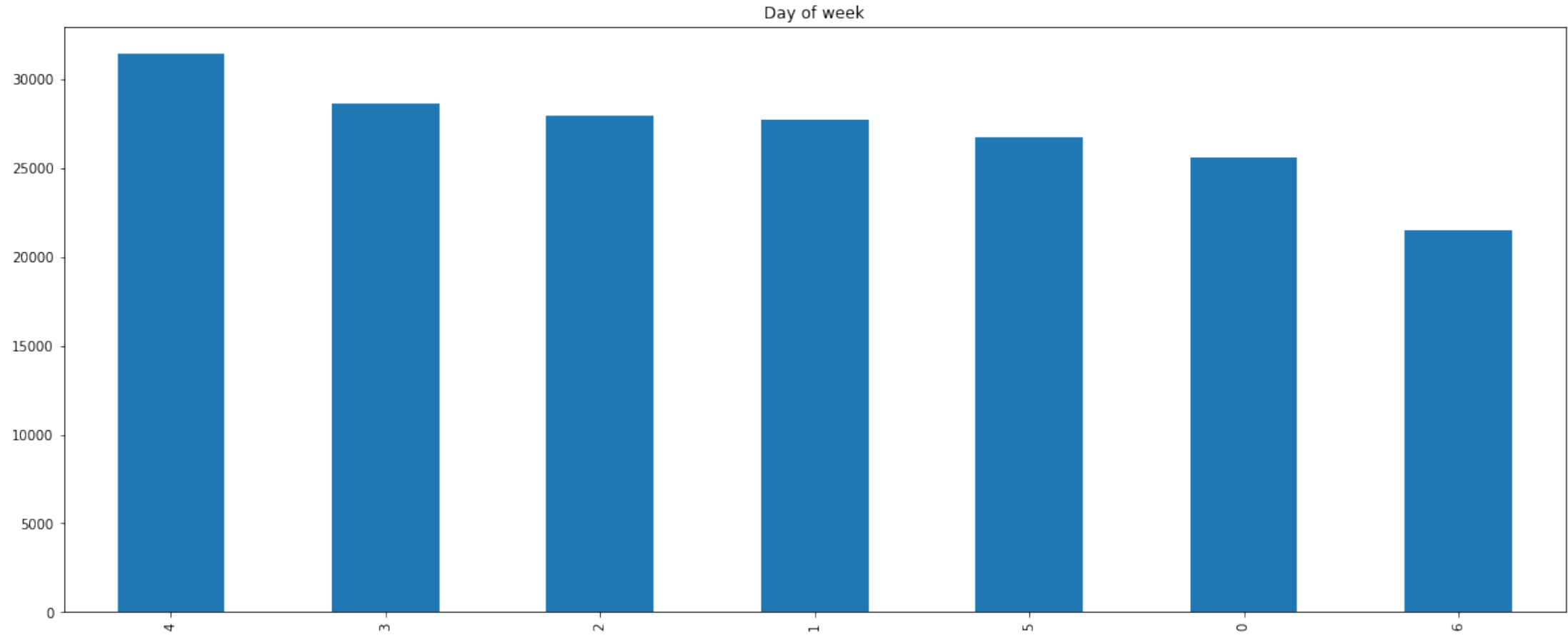
# Distribution of accidents plotted against the address type.
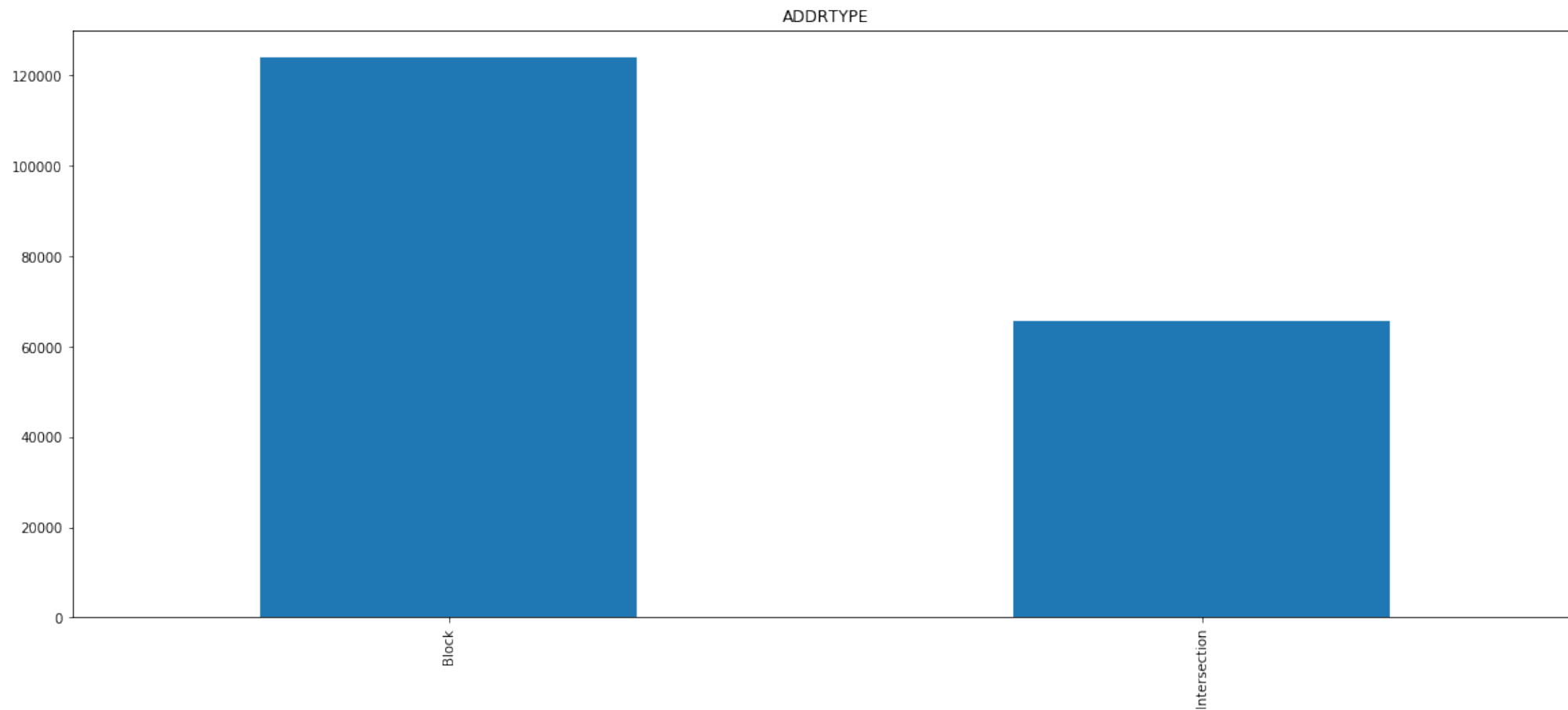
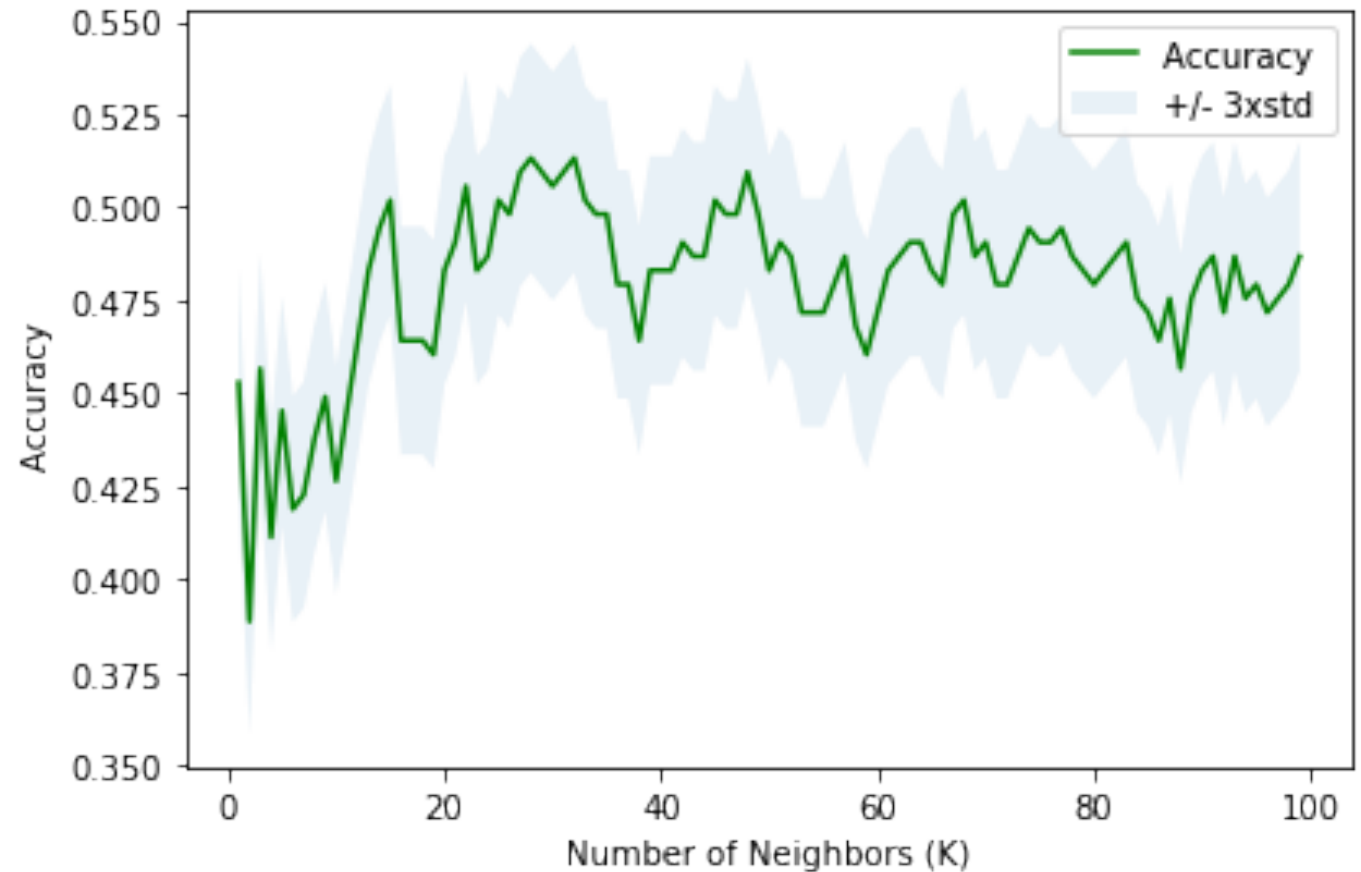# The effect of environmental conditions on road accidents.

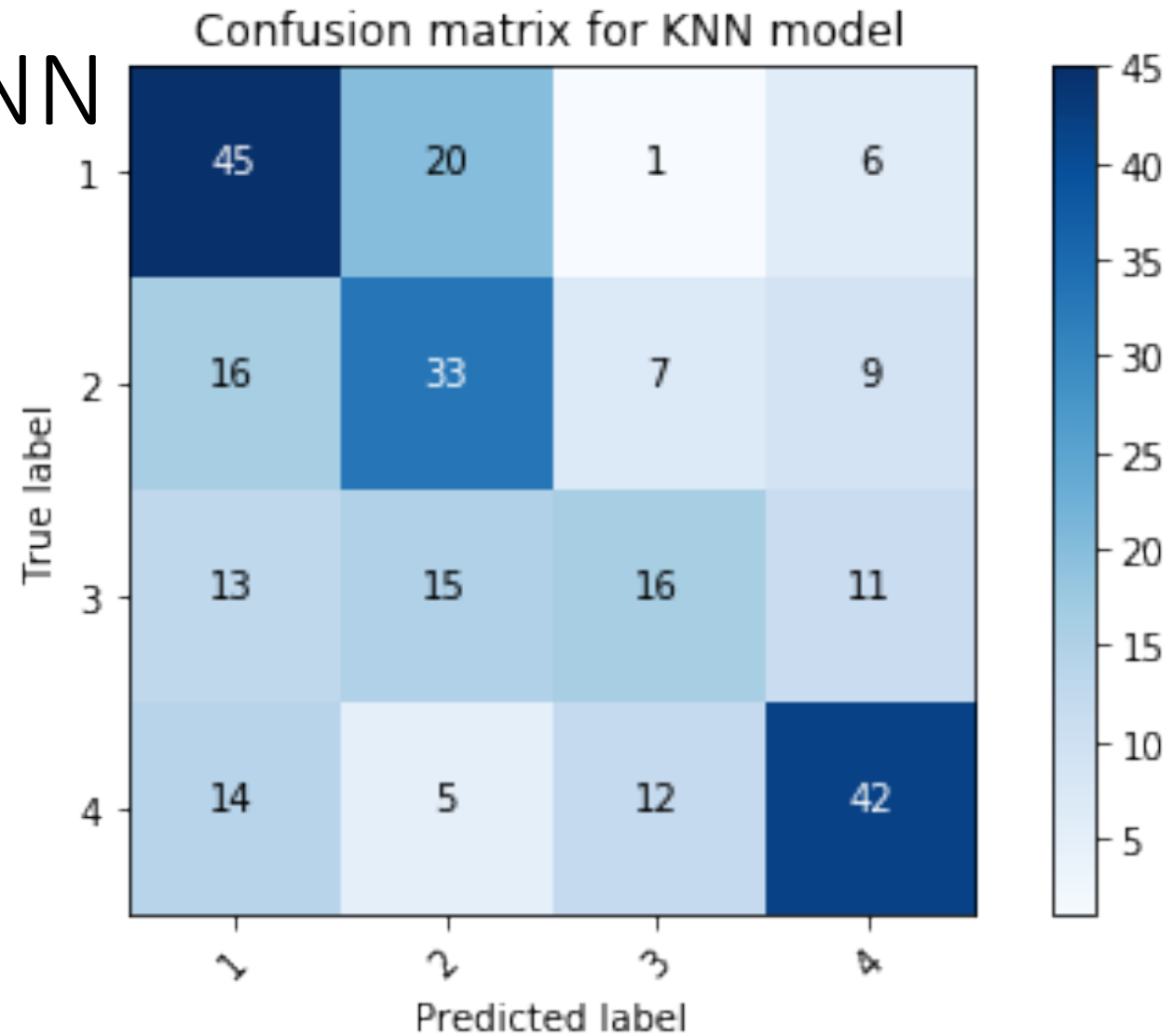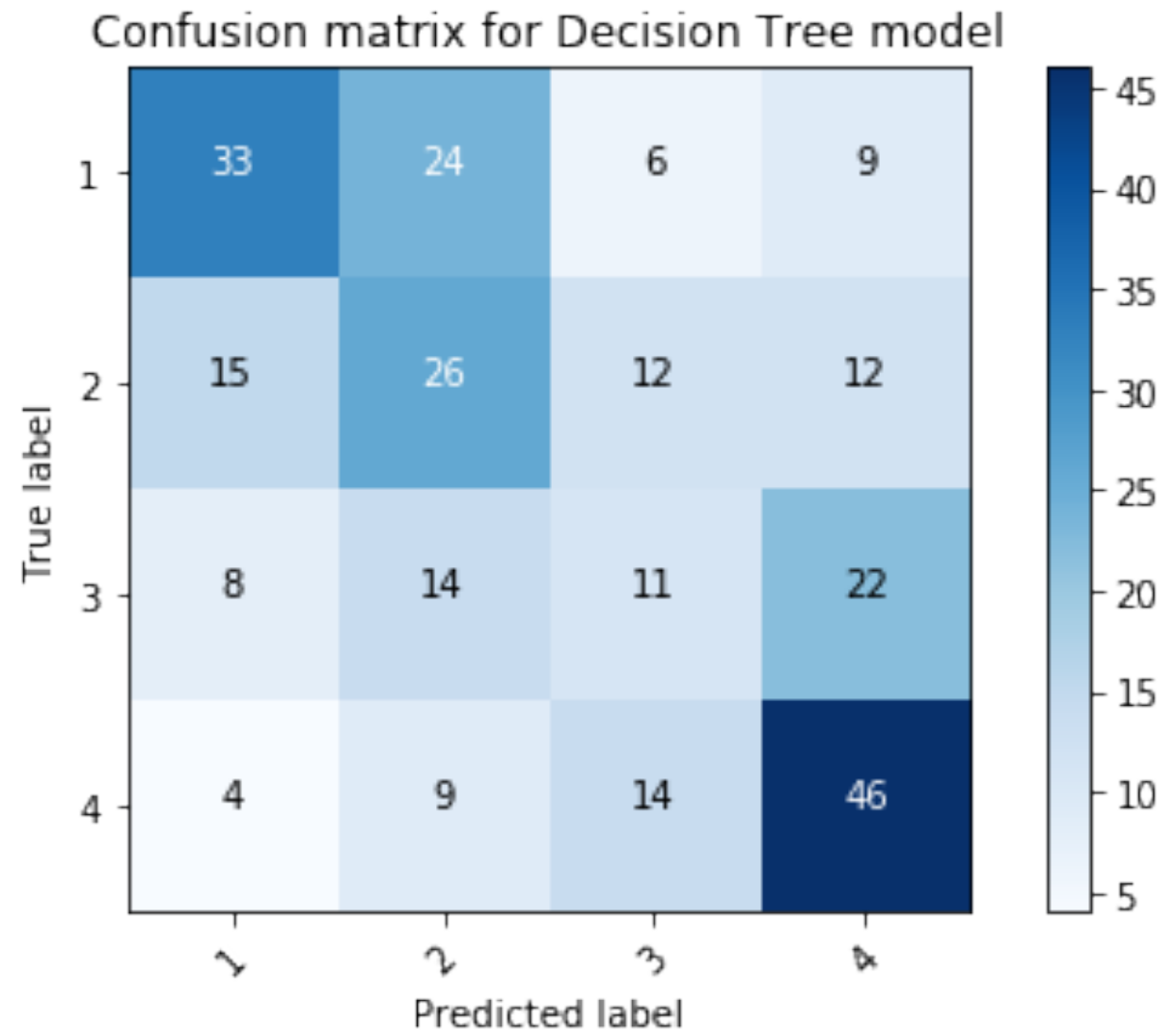# Number of accidents based on personal behaviors.

# Predictive models

- 1. KNN Model



- The best accuracy was with 0.5132075471698113 with k= 28

# Confusion matrix for KNN



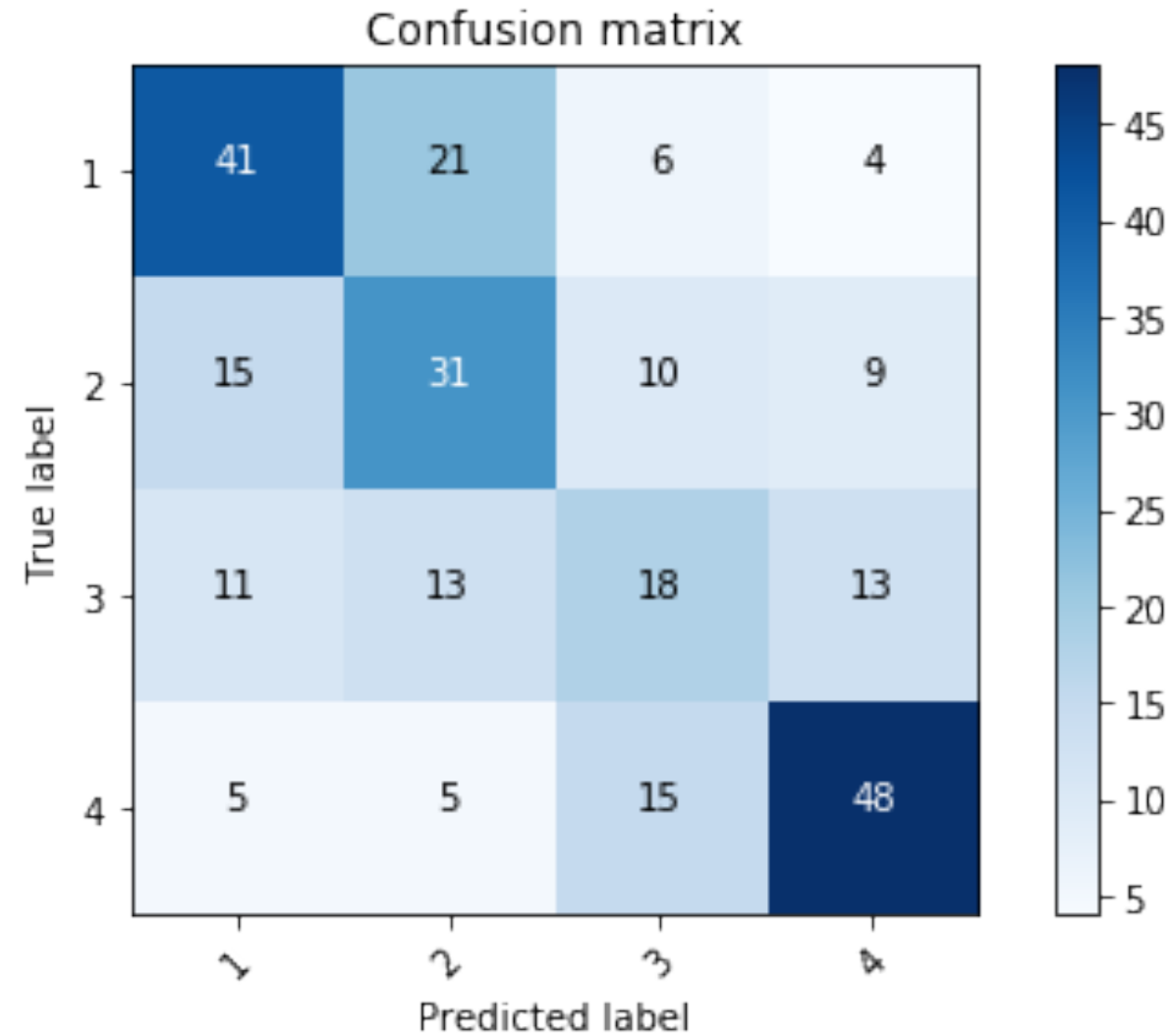Confusion matrix for KNN model

The Jaccard score and F1-score for KNN model are 0.513 and 0.507, respectively.

- 2. Decision Tree Model



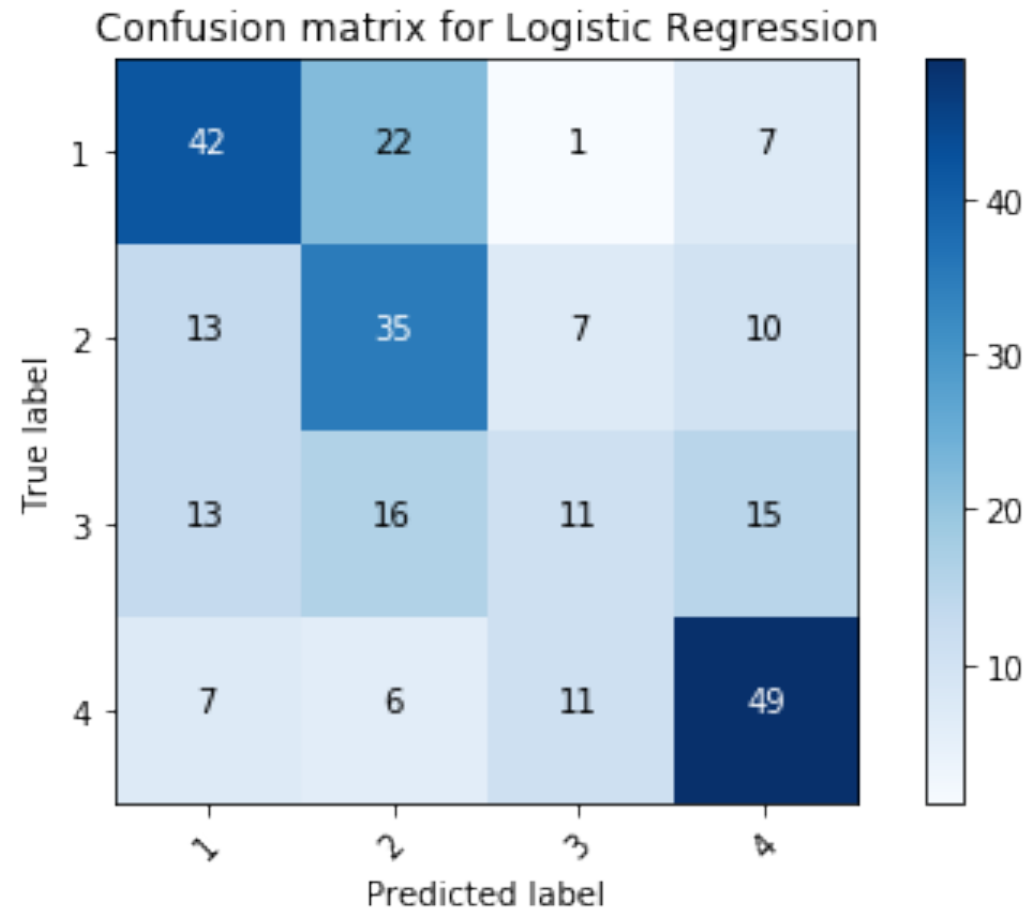Confusion matrix for Decision Tree model

- The Jaccard accuracy and F1 score are 0.44 and 0.43, respectively

- SVM



Confusion matrix

- The model predicted Jaccard accuracy and F1 score are 0.52 and 0.519, respectively.

- Logistic regression



Confusion matrix for Logistic Regression

- The predicted Jaccard accuracy, F1 score, and logloss were 0.52, 0.51 and 1.17, respectively.

# Model Evaluation

| Algorithm | Jaccard | F1-score | LogLoss |
|---|---|---|---|
| KNN | 0.513208 | 0.507234 | NA |
| Decision Tree | 0.437736 | 0.431307 | NA |
| SVM | 0.520755 | 0.519109 | NA |
| LogisticRegression | 0.516981 | 0.503509 | 1.14798 |

# Summary and Conclusion

- In this study, accidents recorded by the Seattle Police Department were used to predict severity of accident when happened.

- Used multiclass classification in machine learning algorithms such as K-nearest neighbor, decision tree, Support Vector Machine and Logistic Regression.

- The model accuracies were evaluated using Jaccard accuracy, F1-score and LogLoss.

- Decision Tree model predicts poorly compared to the other models with ~43% accuracy.

- KNN, SVM, and Logistic Regression models were able to obtain approximately 52% of accuracy.