Applied Data Science Capstone Project- Car accident severity

Sergio Esposito
September 8th, 2020

Can car accident severity be predicted?

Machine learning techniques can be used to predict car accident severity, based on variables like weather and road conditions

Who can be interested on this prediction?

- Insurance companies
- Transport/road traffic government agencies
- Car drivers

Data acquisition and improvement

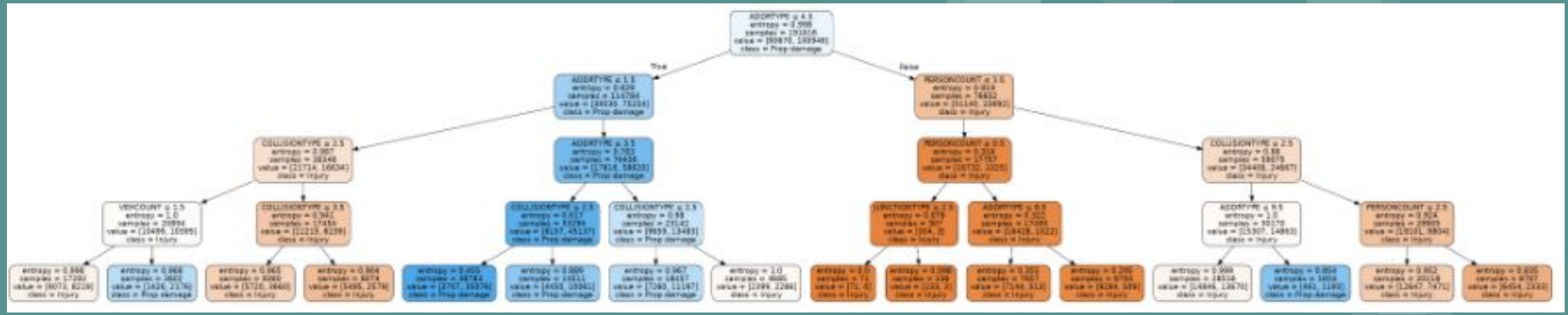
- Dataset provided by Coursera which compiles car accidents in the city of Seattle
  https://drive.google.com/file/d/1uPOlUo2qaF-WYOYq5UljADLEqw1GGfwM/view?usp=sharing

- Variable to predict: SEVERITYCODE
- Attributes: ADDRTYPE, COLLISIONTYPE , PERSONCOUNT, VEHCOUNT, JUNCTIONTYPE, WEATHER, ROADCOND, LIGHTCOND, ST_COLCODE
- Data improvement techniques: clean rows with NaN values, convert object attributes to numeric values, convert the variable to predict to descriptive values, and balance the dataset.

Methodology

- The variable to predict has a discrete set of values → Classification techniques
- Subset of 20,000 rows
  - Logistic Regression: jaccard_similarity_score=0.68625
  - SVM: jaccard_similarity_score=0.73475
  - Decision Tree: accuracy_score=0.7465
  - SVM is around 10 times slower than decision tree
- Chosen technique: Decision Tree

Results



This decision tree has predicted the severity of a car accident with a 69.3% of accuracy

Discussion

- This line of research can be continued by trying other classification techniques like Random Forest and Gradient Boost.
- Voting Ensemble could be used to combine the results of different models.

Conclusion

- With an accurate dataset as input, a Machine Learning model can predict the severity of car accidents.
- This model can be useful to save lives, improve transportation planning, and smart use of financial and infrastructure resources