Car Accident Analysis

Jonathan Graham and Noah Shoap

Objective

- Analyze what factors contribute to severity of car accidents

- Questions:

- Are these factors intuitive?

- Are there any unexpected factors?

- Can these factors accurately predict accident severity?

Car Accident Data

- We acquired a data set from Kaggle which consisted of information about car accidents.

- Among this information was an integer representation of the accident severity

- And a large collection of features, such as weather, location, time, etc.

- The data set was very large, so we trimmed it down to around 20,000 entries.
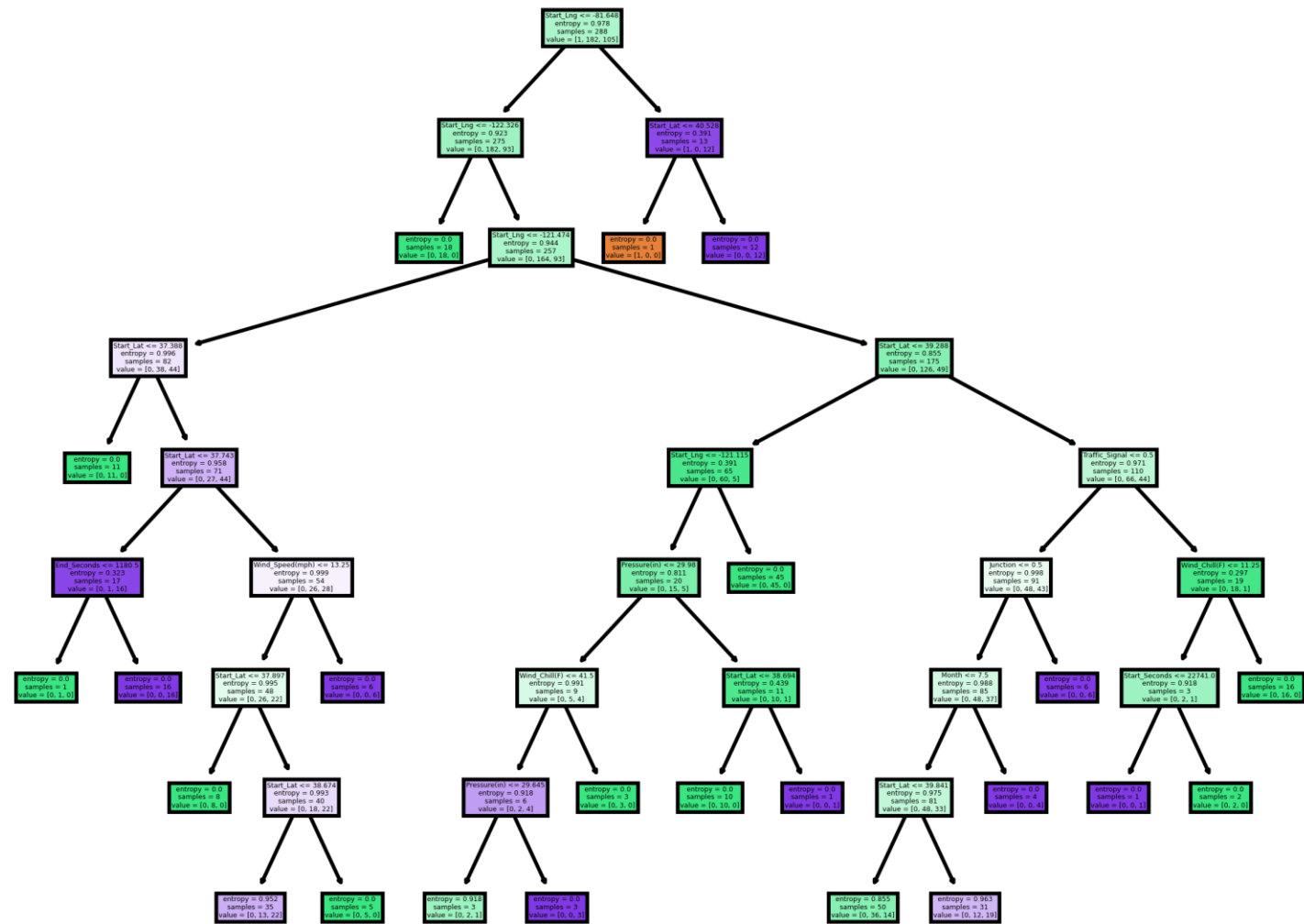
- We did this using Pandas.

Decision Tree

- We opted to use a Decision Tree classifier

- We chose this because we found it simple and intuitive to use

- Fairly accurate (80% accuracy on predictions)

Decision Tree Hyper-parameters

- We tested the tree with various maximum depths:
  - 1
  - 2
  - 4
  - 8
  - 16
  - 32
  - 64
  - 128
  - 256

Best Performing: Max_Depth=8

Results

- This decision tree predicted the severity correctly 83.56% of the time.
- Referencing the splits of the decision tree allowed us to determine what features were most important.
  - Longitude
  - Latitude
  - Traffic Features
    - Signal Light
    - Junction
  - Weather Conditions
    - Wind Speed
    - Pressure
    - Wind Chill
  - Time
    - Month
    - Time of Day

Results Continued

- We found some of these features to be intuitive and expected, such as weather conditions.

- Unintuitively, the longitude and latitude features were the most important splits in the tree. We assume this is likely because of population density in certain areas of the country.

Results Continued

- Questions (and answers):
- Are these factors intuitive?
  - Yes, we found that most of the factors, such as weather and traffic information, were intuitive and we expected to see these used by the Decision Tree.
- Are there any unexpected factors?
  - Yes. We did not expect latitude and longitude to be the most important factors going into our analysis.
- Can these factors accurately predict accident severity?
  - Yes. Our model predicted accurately 83.56% of the time.

Challenges

- The initial dataset required a lot of trimming to be functional. Neither of our computers could open the initial dataset because of how large it was and how many features it had. Ideally, we could have used more entries.

Future Work

- We were content with the accuracy of the decision tree, but using different models and comparing results would be interesting.

- While we suspect that the latitude and longitude features were the most important feature because of population density, that needs to be confirmed. If our assumptions are correct, we could adjust our work to account for population density differences.

Thanks for listening!