Predicting the average amount of car accidents based on current weather conditions, with machine learning techniques

authors: Sigurdur Haukur, Gunnar, Tomas

Introduction

- research question
  - Can the average amount of car accidents be predicted based on current weather conditions?
- Background and importance of predicting car accidents based on weather conditions
  - Car accidents are a major cause of death and injury worldwide
  - Weather conditions are a major factor in car accidents
  - risk of car accidents can be reduced by predicting them based on weather conditions
    * e.g. by warning drivers of dangerous conditions
- Objectives of the study
  - Predict the average amount of car accidents based on weather conditions
  - Determine the relationship between weather conditions and car accidents
  - Determine the predictive capabilities of the model
- scope
  - The study is limited to the relationship between temperature, wind speed, and car accidents
  - The study is limited to the geographical region of Iceland and the USA
  - The study is limited to the time period of 2009 to 2022

Methodology

- Data collection
  - Iceland Dataset (Temperature)
    * weather data from the Icelandic Meteorological Office
      · only historical weather data available consists of **monthly** temperature and rainfall
      · hence, the average amount of car accidents per month was used
      · Average temperature: 4.3893
      · Standard deviation: 4.1323
      · took the average of every weather station in Iceland (261 stations)
      · not enough weather stations had rainfall data, so it was not

used
* car accident data from the Icelandic Road and Coastal Administration
  · Average accidents per month: 246
  · few data points, from 2009 to 2022
  · machine learning models require a large amount of data
  · hence, the model was asked to predict the average amount of car accidents per month for the whole of Iceland
- USA Dataset (Temperature and wind speed)
  * more detailed data, containing the temperature and wind speed at the time of each car accident
  * car accident data from Kaggle
    · 421.675
    · 2.8 million data points, from 2016 to 2021
    · 49 states
- Linear Regression
  - Modeling the relationship between temperature and car accidents (Iceland Dataset)
- Random Forest Regression
  - Incorporating temperature and wind speed to predict car accidents (USA Dataset)

Results and Analysis

- Linear Regression Results (Iceland Dataset)
  - significant relationship between temperature and car accidents
  - too little data, with too much variance(noise)
  - However, the model was able to predict the average amount of car accidents per month for the whole of Iceland

Metric	Value	Interpretation
Mean Squared Error (MSE)	2128.68	Lower values indicate better fit
Mean Absolute Error (MAE)	39.46	Lower values indicate better fit
R-squared Score (R2)	0.14	Higher values indicate better fit
Cross-Validation Scores	[-0.47525246, -1.15262578, 0.17926531, -0.14678218, -0.29697801]	Higher values indicate better fit

- Random Forest Regression Results (USA Dataset)
  - Consideration of temperature and wind speed
    * significant relationship between temperature and car accidents
    * not significant relationship between wind speed and car accidents

for the whole of the USA, since not all accidents are affected by wind speed, and the fact that most accidents occur in urban areas, where wind speed is not a factor
- The model was able to predict the average amount of car accidents per month for the whole of the USA
  * with validation score of .963

Metric	Value	Interpretation
Mean Squared Error (MSE)	394824074.01	lower values indicate better fit
Mean Absolute Error (MAE)	10451.87	lower values indicate better fit
R-squared Score (R2)	-0.38	Higher values indicate better fit

Discussion

- Limitations of the study
  - Neglecting other potential variables affecting car accidents
    * such as rainfall, snowfall, road conditions, etc.
  - Limited amount of data, for the scope of Iceland
  - Limited representation of global patterns in the datasets
- Future Work
  - Inclusion of additional variables
  - Expansion to different geographical regions and diverse weather conditions

Conclusion

- Summary of findings
  - The model was able to predict the average amount of car accidents per month for the whole of Iceland
  - The model trained on the USA dataset was able to predict the average amount of car accidents per month for the whole of the USA based on temperature and wind speed, which suggests that if more data was available, similar models would be able to predict the average amount of car accidents per month for the whole of Iceland based on temperature and wind speed
- Implications and recommendations for further research
  - the ability to predict the average amount of car accidents based on weather conditions can be used to reduce the risk of car accidents
    * e.g. by warning drivers of dangerous conditions
  - further research should be done on the relationship between weather conditions and car accidents
    * e.g. by including additional variables, such as rainfall, snowfall, road conditions, visibility etc.

∗ further data collection is required for this