

Big Data - Deliverable 1

Group 21

1. Team:

a. Members:

1. Lakshmi Tulasi Bhavanam (801275438)
2. Ram Deepak Vinjam (801275255)
3. Jaswanth Tumati (801274978)
4. Rohith Sri Sai Mukkamala (801275202)
5. Jaswanth Yadla (801275432)

b. Communication plan to include project artifact repository

- (1) We will be having Zoom calls every week.
- (2) Github repository- https://github.com/developer-rohith/bigdata_project for discussing the project progress.

2. Business Problem:

Around the world, reducing road accidents is a significant concern for public safety. 1.25 million people died in traffic accidents worldwide in 2013, up from 68 countries in 2010 according to a global progress report on traffic safety.

To make the roads safer, accident prediction is crucial for public transportation optimization, permitting better routes, and cost-effectively enhancing the transportation infrastructure. Due to its importance, accident prediction and analysis has generated a lot of study in recent years.

To a finest of our knowledge, US-Accidents, which contains over 2.25 million records of traffic accidents gathered for the entire United States over three years, is the first nationwide dataset of this size. We were able to gain a range of insights about the place, timing, climate, and comments point of an accident from this dataset. We think that US-Accidents offers a context for upcoming studies on the study and forecasting of transportation accidents. Regarding our own project study, we intend to use this dataset to carry out real-time traffic accident prediction.

3. Dataset:

Traffic Accident Dataset: This dataset on car accidents in the USA includes data from all 49 states. Several APIs that offer live traffic collision data were used to gather the accident data, which covered the period between February 2016 to December 2021. The federal and state dept of transportation, police organizations, license plate readers, and traffic sensors embedded in the road networks are only a few of the organizations whose traffic data is broadcast through these APIs.

Data schema: Few data attributes are given here.

1. ID: This serves as the accident record's special identifier.
2. Severity: shows the accident's severity as a number ranging from one and 4.
3. Weather_Timestamp: displays the weather observation record's time stamp (in local time).
4. Temperature(F): display of temperature (in Fahrenheit).
5. Wind_Chill(F): display of wind chill (in Fahrenheit).
6. Wind_Direction: Direction of wind
7. Wind_Speed(mph): Speed of the wind in miles per hour
8. Precipitation(in): Precipitation amount in inches
9. Weather_Condition: demonstrates the weather condition like rain, thunderstorm, snow etc.
10. Traffic_Signal: a POI annotation that denotes the presence of a traffic signal in the area.

4. Research Objectives and Questions:

1. Predictions:
 - a. Prediction of accident during different weather conditions in real time.
 - b. Hotspots where the accidents are more happening.
2. Analyze the cause and effect of the accidents
 - a. Identifying the main root cause of the accident such as traffic signal, weather conditions or road conditions.
 - b. Understanding the influence of weather or other environmental factors on the likelihood of accidents.