

CMP4011 Big Data and Cloud Computing

Project Report

Team 10

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Sec** | **B.N** | **Code** |
| Ahmed Osama Helmy | 1 | 5 | 9213061 |
| Omar Mahmoud | 1 | 29 | 9210758 |
| Abdallah Ahmed | 1 | 25 | 9210652 |
| Aliaa Gheis | 1 | 27 | 9210694 |

# Descriptive Insights

**1. Weather Condition Severity Correlation**

* *Insight*: Analyze how specific Weather\_Condition values correlate with accident Severity levels.
* *Algorithm*: KMeans clustering on weather parameters (Temperature, Humidity, Visibility, Wind\_Speed, Precipitation) to identify dangerous weather profiles.

**2. Geographical Accident Hotspot Identification**

* *Insight*: Use Start\_Lat, Start\_Lng, End\_Lat, and End\_Lng to identify areas with concentrated accident occurrences.
* *Algorithm*: Clustering to discover spatial accident hotspots across different States and Counties.

**3. Visibility Impact Assessment**

* *Insight*: Analyze how the Visibility(mi) field correlates with accident Severity and Distance(mi).
* *Algorithm*: Decision tree analysis to identify critical visibility thresholds that predict severe accidents.

**4. Traffic Signal Effectiveness**

* *Insight*: Compare accident Severity and frequency in locations with Traffic\_Signal = True versus False.
* *Algorithm*: Support Vector Machines (SVM) to classify the effectiveness of traffic signals in preventing severe accidents.

**5. Day/Night Accident Pattern Analysis**

* *Insight*: Examine how Sunrise\_Sunset, Civil\_Twilight, Nautical\_Twilight, and Astronomical\_Twilight values affect accident patterns.
* *Algorithm*: Random Forest classification to identify features that distinguish day versus night accidents.

**6. Road Feature Impact Analysis**

* *Insight*: Measure how road features (Junction, Crossing, Roundabout, Stop, Bump, etc.) impact accident Severity.
* *Algorithm*: Gradient Boosted Trees to rank the importance of road features in determining accident severity.

**7. Precipitation-Based Risk Assessment**

* *Insight*: Analyze how Precipitation(in) values and Weather\_Condition correlate with accident frequency and Severity.
* *Algorithm*: Logistic regression to quantify precipitation thresholds that significantly increase accident risk.

# Predictive Insights

**1. Weather-Based Accident Prediction**

* *Insight*: Predict accident probability based on Temperature, Wind\_Chill, Humidity, Pressure, Visibility, Wind\_Direction, Wind\_Speed, and Precipitation.
* *Algorithm*: Random Forest regression to model how weather parameter combinations affect accident likelihood.

**2. Accident Duration Forecasting**

* *Insight*: Predict the duration (End\_Time minus Start\_Time) of accidents based on location, weather, and road features.
* *Algorithm*: Gradient Boosting regression to forecast how long traffic will be affected by various accident types.

**3. Severity Prediction Model**

* *Insight*: Build a model to predict accident Severity levels using all available environmental and infrastructural features.
* *Algorithm*: Logistic regression to classify severity categories from multiple input features.

**4. Location-Based Risk Scoring**

* *Insight*: Create risk scores for different geographical areas based on historical accident patterns.
* *Algorithm*: KNN to identify similar location characteristics across the dataset and assign risk scores.

**5. Weather Threshold Identification**

* *Insight*: Determine critical thresholds for Temperature, Visibility, Wind\_Speed, and Precipitation where accident risk significantly increases.
* *Algorithm*: Decision trees to identify weather breakpoints.

**6. City and County Risk Profiling**

* *Insight*: Build risk profiles for Cities and Counties based on their historical accident patterns.
* *Algorithm*: Random Forest to identify location-specific risk factors.

**7. Road Feature Optimization Recommendations**

* *Insight*: Predict which road feature additions (Traffic\_Signal, Traffic\_Calming, etc.) would most effectively reduce accident risk in specific locations.
* *Algorithm*: Trying different predictive models