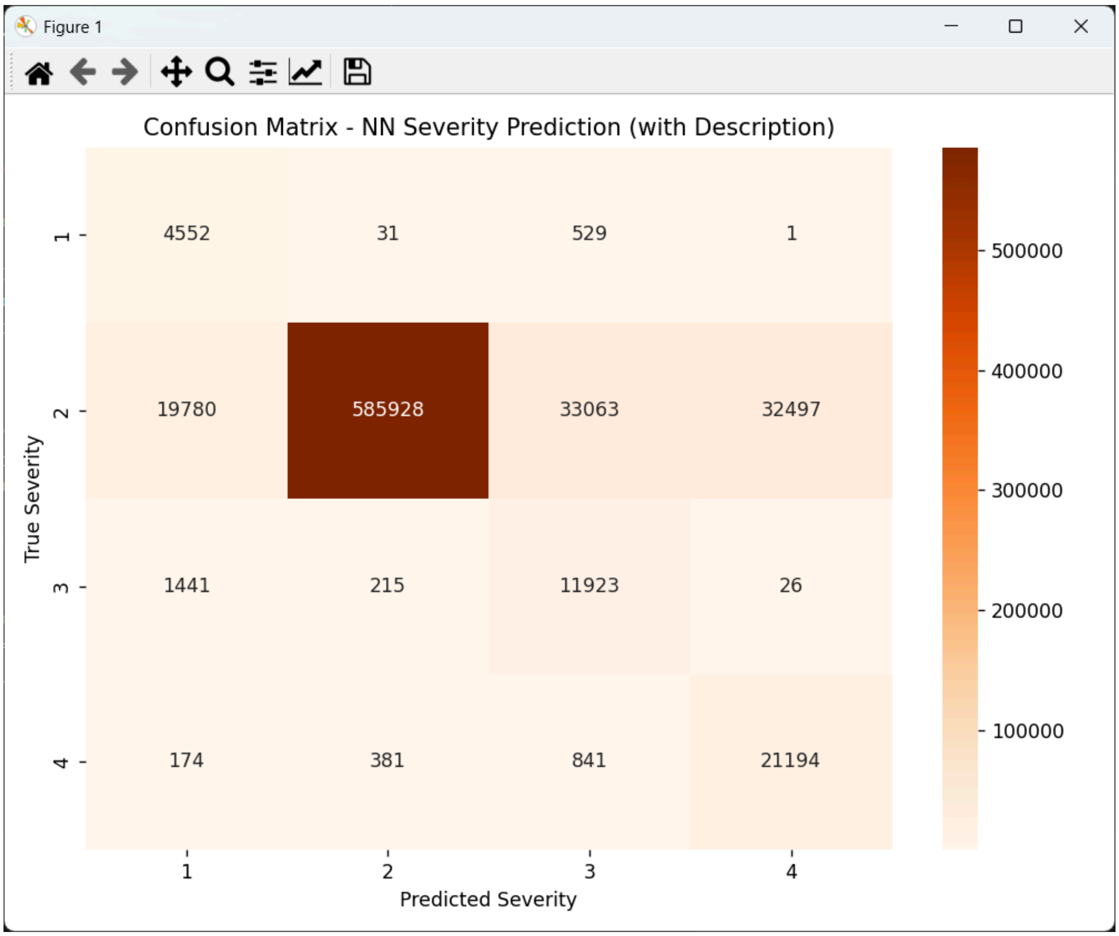
Visualization Results

Shervan Shahparnia and Nathan Cohn - Group 1

Visualization 1: Confusion Matrix - Neural Network Severity Prediction



Message:

This confusion matrix visualizes the performance of a neural network model in predicting traffic accident severity (Levels 1 to 4), using weather and textual description-based features.

Key Insights:

- The model performs best at predicting Severity 2, with 585,928 correct predictions.
- However, there's noticeable confusion between Severity 2 and other classes, especially:

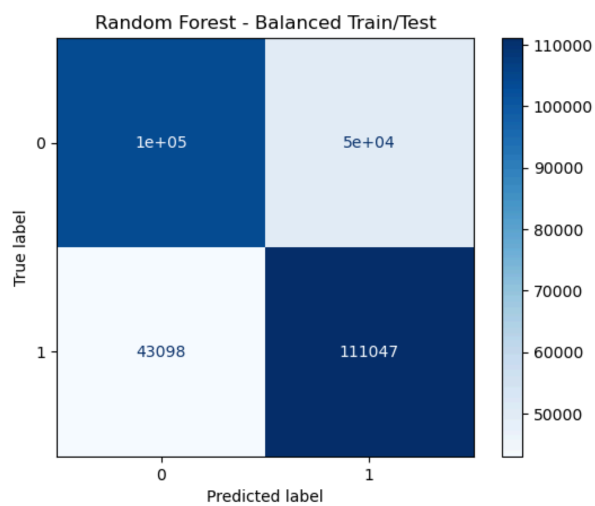
    - Many actual Severity 2 cases are misclassified as Severity 3 (33,063) or Severity 4 (32,497).
    - A sizable number of Severity 3 and Severity 4 cases are misclassified as Severity 2.

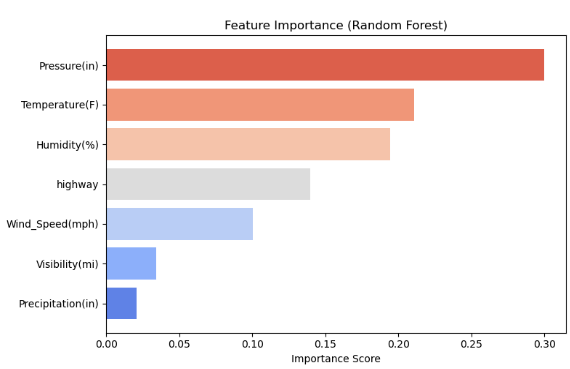
- Severity 1 and Severity 4 are much less frequent and harder to detect accurately due to class imbalance.

Effectiveness:

- The use of a heatmap clearly emphasizes areas where the model is most confident (dark brown for Severity 2).
- It also reveals where the model struggles, such as misclassifying higher severity levels as less severe ones.
- This visualization effectively communicates both strengths and weaknesses of the model across all severity classes.

Visualization 2: Confusion Matrix - Random Forest Model (Balanced Train/Test)





Message:

This confusion matrix illustrates the performance of a Random Forest classifier trained and tested on a balanced dataset for binary accident severity prediction:

- 0 = Low severity (Levels 1–2)
- 1 = High severity (Levels 3–4)

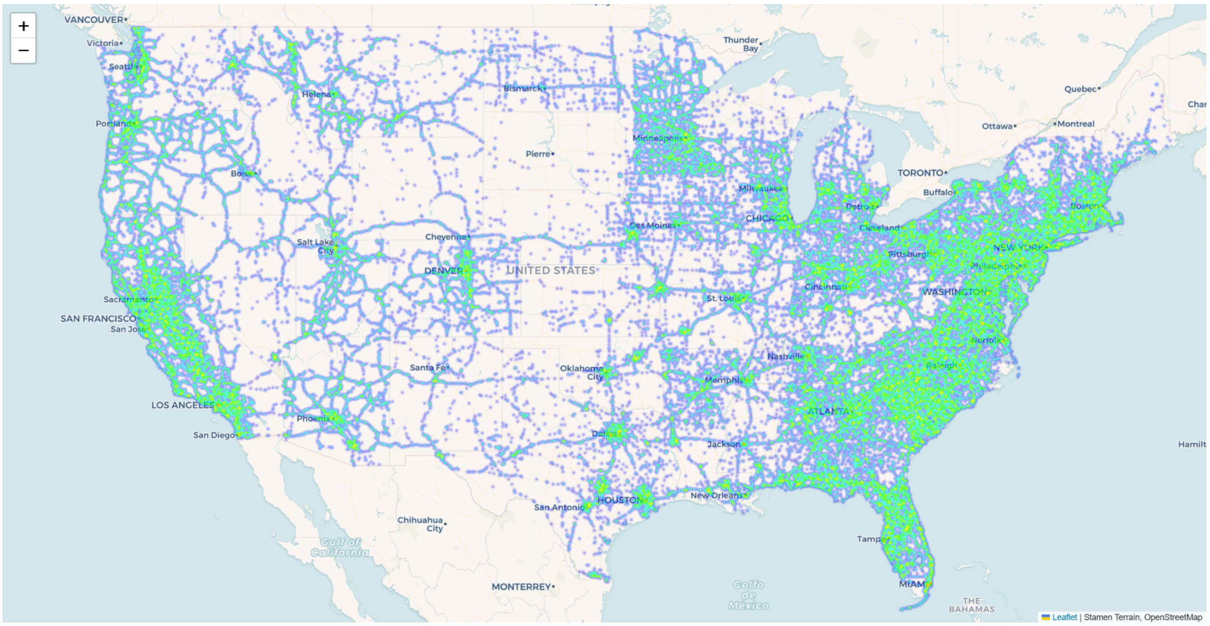
Key Insights:

- The model correctly predicted 111,047 high-severity accidents, showing strong performance in identifying dangerous incidents.
- 100,000 low-severity accidents were also correctly predicted.
- Some misclassification remains:
    - 43,098 high-severity cases were incorrectly predicted as low (false negatives).
    - 50,000 low-severity cases were predicted as high (false positives).
- The model achieves a good balance between precision and recall, which is essential for fair evaluation when both classes matter (e.g., emergency prioritization).

Effectiveness:

- The balanced dataset ensures that the model is not biased toward the majority class.
- The color shading and numerical values help clearly convey the true positives vs. false predictions.
- This visualization effectively communicates the impact of balancing data on model fairness and performance.

Visualization 3: U.S. Traffic Accident Heatmap (Geospatial Distribution)



Message:

This interactive heatmap visualizes the geographic distribution of traffic accidents across the United States. It uses a sampled set of 3 million data points from the original dataset. Sampling was applied to improve clarity and performance, especially when viewing the map at a nationwide scale.

Key Insights:

- Accident density is highest along urban corridors and interstate highways, forming a clear outline of the U.S. road network.
- Areas with the most accident activity include:
    - The East Coast corridor from Boston to Miami
    - The West Coast, especially around Los Angeles, San Francisco, and Seattle
    - The Midwest, with noticeable concentrations near Chicago, Minneapolis, and Detroit
    - Major metro regions in the South, such as Houston, Dallas, and Atlanta

- In contrast, rural and central parts of the country show fewer incidents, which aligns with lower population and traffic levels.

Effectiveness:

- Reducing the map to a sample of 3 million points makes it more readable while still showing the overall traffic pattern. It also allows for faster loading and better observation of the map while zooming and panning.

- The color gradient (green to blue to purple) shows areas of high and low accident concentration.

- This visualization clearly communicates spatial risk trends and helps identify areas that may benefit from traffic safety improvements, urban planning, or targeted policy decisions.