**Probability Density Functions (PDFs)**

A graph with a green line

Description automatically generatedA graph with a blue line

Description automatically generated

A graph of a function

Description automatically generated

The probability density functions (PDFs) for x\_mean, y\_mean, and z\_mean reveal distinct distribution patterns that highlight unique characteristics of the dataset. The PDF for x\_mean shows a sharp, symmetric peak around 0, suggesting a normal distribution with low variance. This indicates that most data points for x\_mean are tightly concentrated near the mean, with relatively few extreme values. In contrast, the y\_mean distribution is highly skewed, with almost all density concentrated near 0 and a long tail extending into negative values. The 1e−7 scaling factor emphasizes the sparsity of y\_mean, as the density values are extremely small. This skewness implies an asymmetric distribution or the presence of significant outliers. Similarly, the PDF for z\_mean exhibits a peak near 0 with a long positive tail, but unlike y\_mean, the tail extends toward large positive values. The 1e−7 scaling once again highlights sparse data, with z\_mean showing large outliers in the positive range. Together, these PDFs demonstrate varying patterns of distribution, ranging from symmetric and compact (x\_mean) to highly skewed with pronounced tails (y\_mean and z\_mean). These differences suggest that the dataset’s features vary greatly in their statistical properties, potentially impacting model performance.

**3D Surface Plot**

**A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generatedA graph with a colorful curve

Description automatically generated with medium confidence**

The 3D surface plot captures the interactions between x\_mean, y\_mean, and z\_mean, illustrating their relationships in a visual and quantitative manner. The plot highlights a smooth surface with regions of curvature, reflecting non-linear dependencies between these variables. Along the y\_mean axis, z\_mean exhibits a gradual increase, indicating a potential correlation between these variables. Conversely, z\_mean shows minimal variation along the x\_mean axis, suggesting that x\_mean may have a weaker influence on z\_mean. The plot’s smoothness indicates that the relationships among the variables are continuous, but the large scale of z\_mean (on the order of 1e8) emphasizes substantial disparities in magnitude across the dataset. These disparities could introduce challenges in modeling, as the relationships may be dominated by extreme values of z\_mean. Different views of the surface plot provide further insight into how these variables interact, with curvature indicating non-linear patterns that may require advanced modeling techniques to capture effectively. Overall, the 3D plot provides a clear visualization of how the variables are related, with a particular emphasis on the dominant impact of y\_mean on z\_mean and the relative insignificance of x\_mean in driving variation.

**Neural Network and Performance**

A graph with a line drawn on it

Description automatically generated

The Transformer neural network demonstrated noticeable improvement during training, as evidenced by a reduction in loss from 0.33 to 0.11 over five epochs. This decrease indicates that the model was learning effectively from the training data, refining its internal representations to better predict the target variable. However, the final test accuracy of 13% highlights significant issues with generalization. Despite the reduced training loss, the model struggles to perform on unseen data, suggesting that it has either overfitted to the training data or failed to capture meaningful patterns from the features. Potential causes for the low accuracy of 14% can be attributed to three primary factors: imbalanced dataset, overfitting, and feature quality. First, the dataset shows a significant class imbalance, with 25,772 instances labeled as intoxicated (35.5%) compared to 46,750 labeled as non-intoxicated (64.5%). This imbalance means the model may have learned to prioritize the majority class (non-intoxicated), resulting in poor predictions for the minority class (intoxicated). Models trained on imbalanced datasets often fail to generalize well to the underrepresented class, leading to low overall accuracy in binary classification tasks. Second, the substantial drop in training loss from 0.31 to 0.13 alongside poor test performance suggests overfitting. This implies that while the model has captured patterns in the training data, it has failed to generalize these patterns to unseen data, likely memorizing noise, or redundant features instead of learning the underlying relationships. Unfortunately, we adjusted the dropout rate to prevent overfitting as well as changing the dimensions of the model to 64, 128, 256, 512 and yet we still produced low predictive ability. Lastly, the quality of the input features may be a limiting factor. If some features do not strongly correlate with the target variable (intoxication status), the model may struggle to extract meaningful patterns for prediction. This could result in the Transformer model’s capacity being underutilized or focused on irrelevant signals, ultimately leading to reduced predictive accuracy. Choosing a correlation threshold such as greater than 0.5 or less than -0.5 and only using the variables that cross this threshold as input features to the model could potentially help the predictive prowess.