**Machine Learning Model Training and Evaluation**

This report summarizes the approach, findings, and results of the machine learning project aimed at anomaly detection using different models. The models evaluated include Random Forest, Support Vector Machine (SVM), and Neural Network.

**Data Preprocessing**

Data preprocessing involved encoding categorical variables and scaling numerical features. The dataset was split into training and validation sets to ensure robust evaluation. Key steps included:

* Encoding categorical variables using one-hot encoding.
* Scaling numerical features using StandardScaler.
* Splitting the training data into 80% training and 20% validation sets.

**Feature Selection**

Feature selection was performed using ANOVA F-value to select the top features contributing to the target variable. The top features were visualized to understand their importance.

**Model Training and Evaluation**

Three models were trained and evaluated on the selected features:

* Random Forest
* Support Vector Machine (SVM)
* Neural Network

Each model was assessed based on accuracy, precision, recall, and F1-score. The Random Forest model showed the highest accuracy and robust performance metrics across the board.

**Model Comparison**

The models were compared using various metrics:

* **Random Forest**: Achieved the highest accuracy, precision, recall, and F1-score. It had the best performance in distinguishing between normal and anomalous classes.
* **Support Vector Machine (SVM)**: Performed well but slightly behind Random Forest in terms of accuracy and other metrics.
* **Neural Network**: Showed competitive performance but did not surpass Random Forest. It provided valuable insights into the data despite higher computational cost.

**Feature Importance Analysis**

The feature importance analysis using the Random Forest model highlighted the most significant features contributing to the classifications. This analysis provided deeper insights into the underlying data patterns.

**Conclusion**

Among the models tested, the Random Forest classifier demonstrated outstanding performance with the highest accuracy and robust evaluation metrics across the board. It is recommended as the best-performing model for this anomaly detection task, balancing accuracy, interpretability, and computational efficiency. This project underscores the importance of rigorous model evaluation and comparison to identify the most effective machine learning approach for a given problem. The Random Forest classifier emerged as a reliable and accurate choice, making it suitable for deployment in real-world applications where anomaly detection is critical.