SMS Fraud Detection System - Executive Report

Executive Summary

This report presents the findings of our latest financial fraud analysis using machine learning models on email communication data. The model demonstrates high accuracy in identifying potential spam emails, with an impressive precision rate of 99.5%. However, it's essential to acknowledge a slight discrepancy in recall (93.3%), indicating that there may be instances where genuine emails are being incorrectly classified as spam.

Model Evaluation

The model's performance has been thoroughly evaluated using key metrics such as accuracy, precision, recall, and F1 score. The high precision rate indicates a low false positive rate, which is crucial in our context to minimize potential missed opportunities from legitimate emails. However, the relatively lower recall suggests that efforts should be made to improve the model's sensitivity towards spam emails without compromising its ability to identify genuine emails.

Risk Indicators (lexical + structural)

The analysis of top unigrams in spam emails provides valuable insights into potential risk indicators. The prevalence of terms such as "free," "txt," "mobile," "text," "stop," "claim," "reply," "www," and "prize" are common in spam messages, suggesting that emails containing these words should be scrutinized more closely. Additionally, the higher average character count observed in spam emails could serve as another indicator of potential fraudulent communication.

Trust & Interpretability (e.g., via LIME)

To ensure the trustworthiness and interpretability of our model, we employ techniques such as Local Interpretable Model-agnostic Explanations (LIME).

These methods allow us to understand how specific features influence the model's predictions, thus providing valuable insights into the decision-making process of the machine learning model.

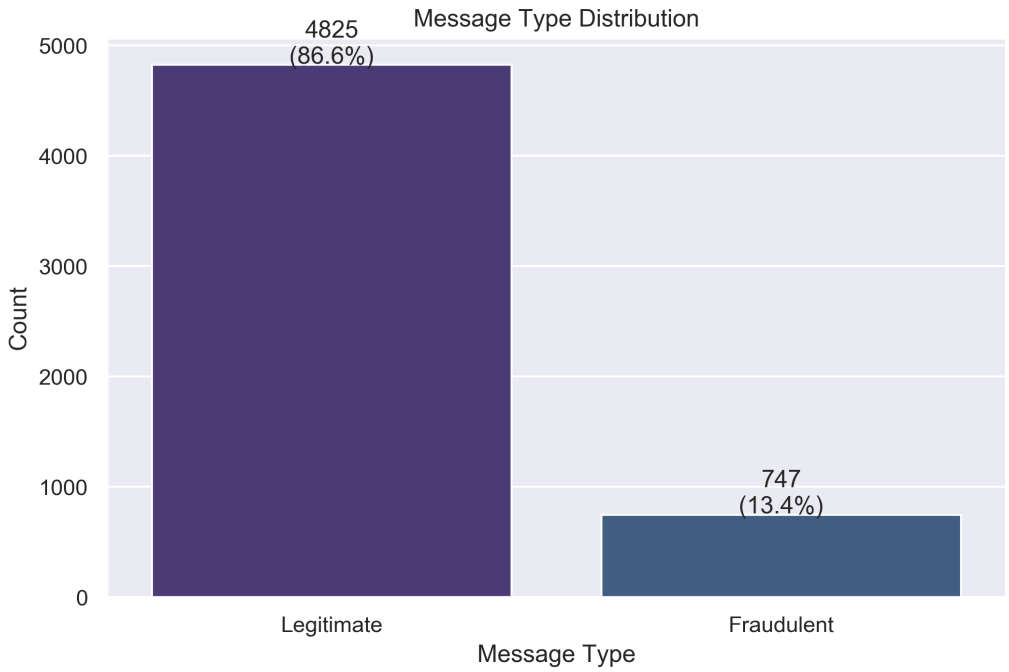
Governance & Recommendations

The governance of our fraud detection system is paramount to its success. Regular audits and reviews of the model's performance should be conducted to ensure it remains effective in evolving threat landscapes. Additionally, a feedback mechanism should be established to continuously improve the model by incorporating user inputs on correctly and incorrectly classified emails.

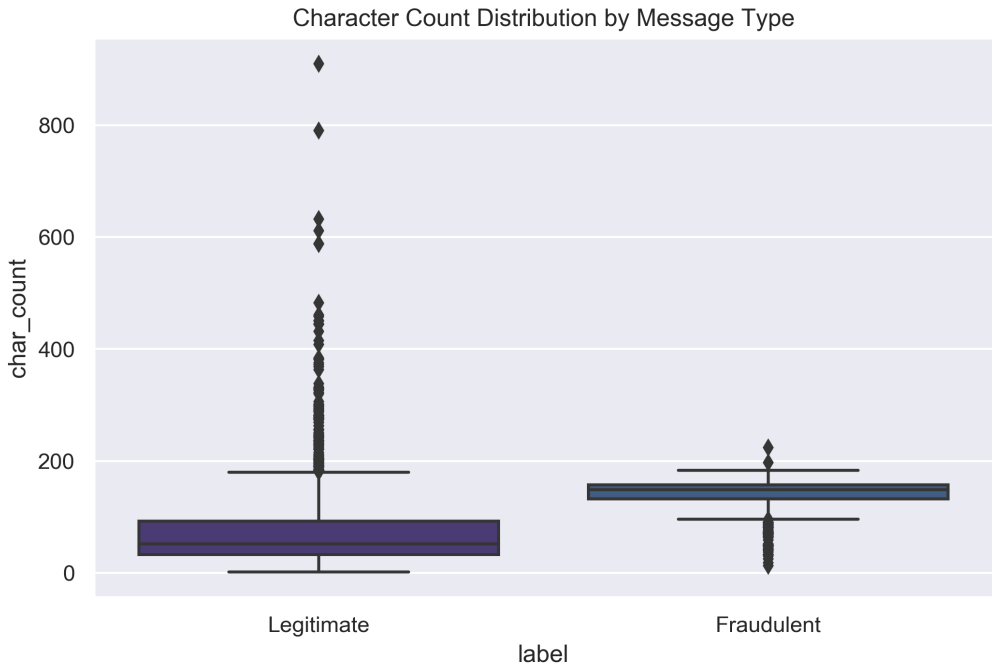
Lastly, it is recommended that the focus should not only be on improving the model's recall rate but also on enhancing its ability to handle large volumes of data while maintaining high performance levels. This will ensure our system remains efficient and effective in safeguarding our organization from financial fraud.

Figures & Visuals

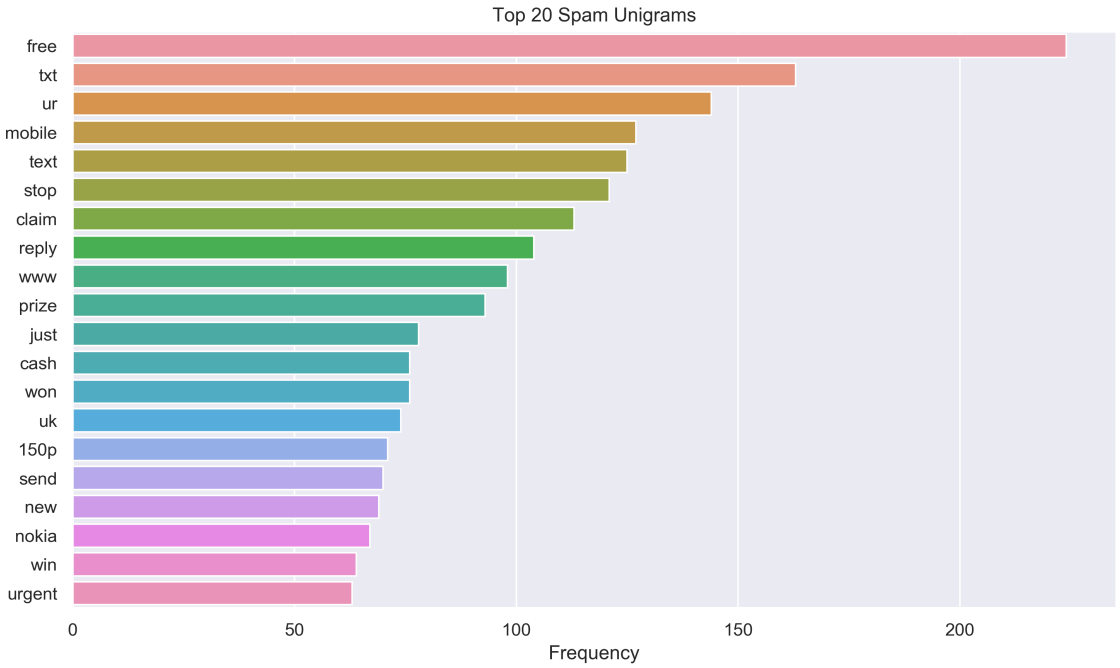
Class Distribution



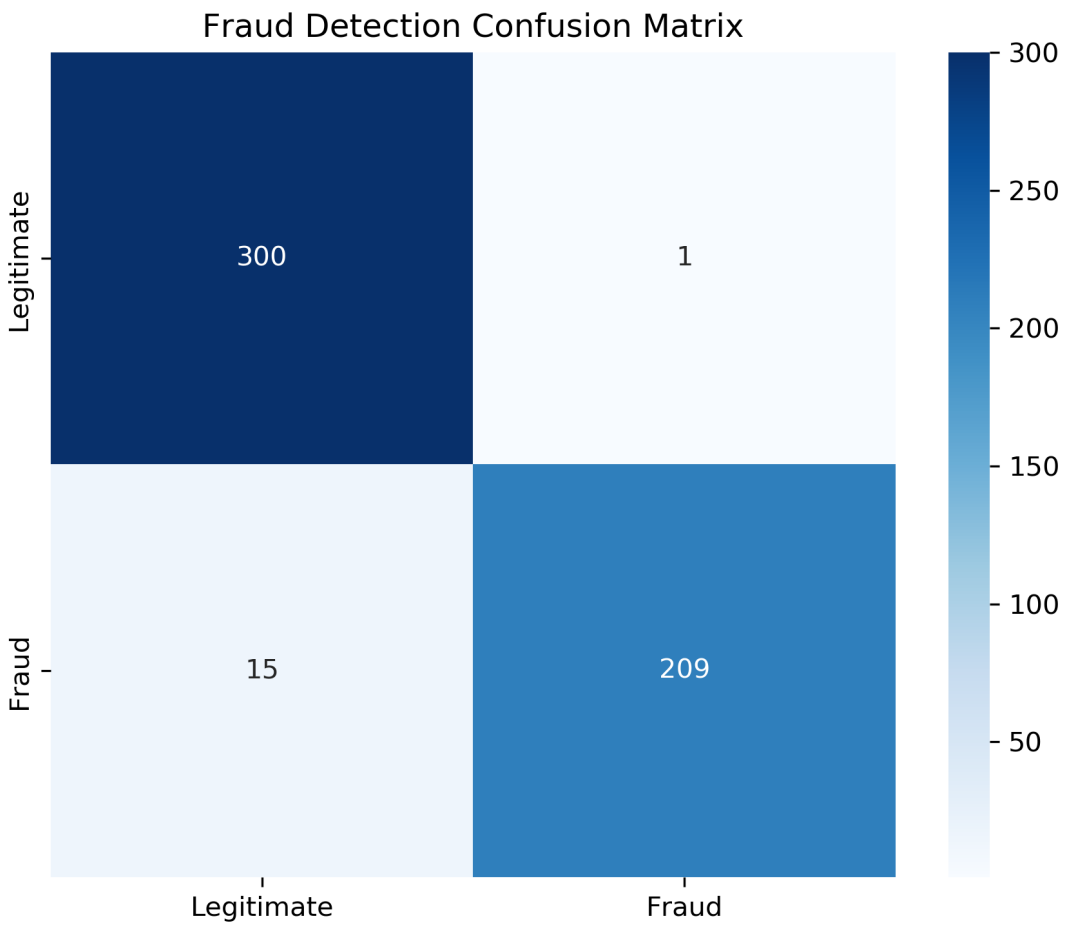
Text Length Distribution



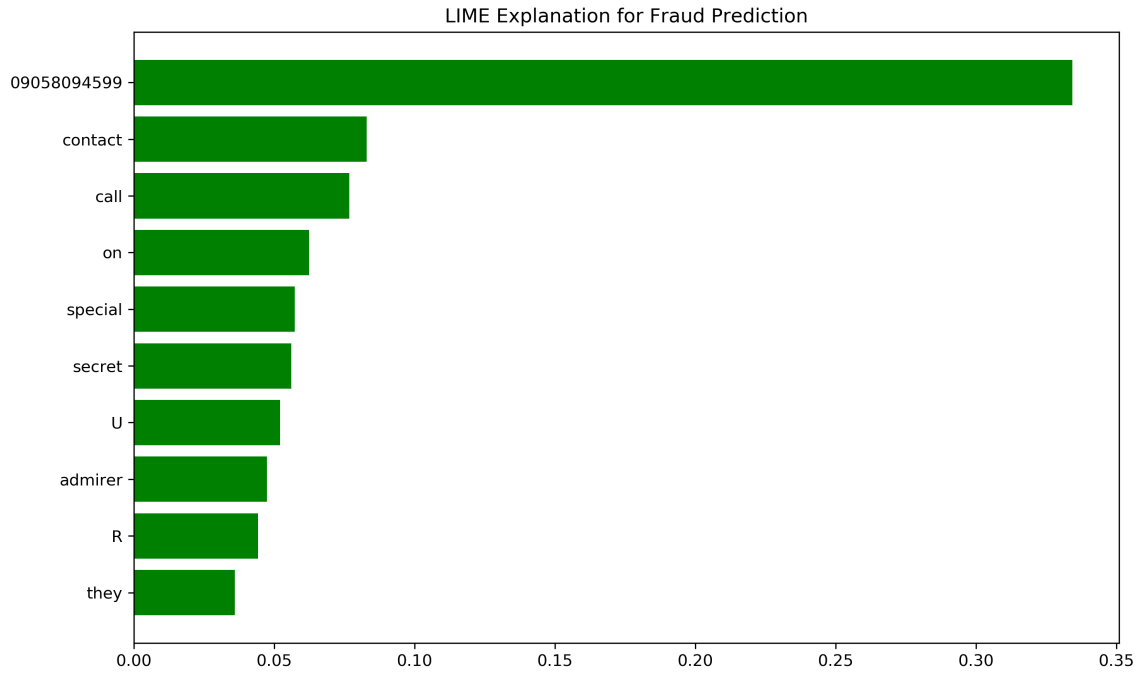
Top Fraudulent Unigrams



Confusion Matrix



LIME Explanation



Generated by AI Analyst · Date: 2025-05-14