Optimizing Email Management for Organizations through Multi-Class Classification Using Llm Models

Vishal Tyagi
Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, India

Sarthak Marwah
Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, India

Dr. S. Nagadevi
Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, India

Abstract

Efficient email management is crucial for organizations due to the growing volume of electronic communication. This research explores multi-class email classification using both traditional and advanced machine learning approaches. A Support Vector Machine (SVM) model serves as the baseline, achieving an accuracy of 67.07% and an F1-score of 62.01%. Key performance metrics for the SVM model include a Mean Squared Error (MSE) of 2.8710, Root Mean Squared Error (RMSE) of 1.6944, Mean Absolute Error (MAE) of 0.9885, Standard Deviation (SD) of 1.6741, Correlation Coefficient (R) of 0.4144, and Coefficient of Determination (R^2) of 0.4320. The model demonstrates moderate performance with acceptable consistency, making it suitable for structured email classification tasks. In comparison, a Large Language Model (LLM) approach based on XLNet is finetuned on the same dataset to leverage contextual embeddings for enhanced classification. The LLM achieves an accuracy of 69.98% ≈ 70% and an F1-score of 54.89%, demonstrating slightly improved accuracy while showing lower F1 performance compared to SVM. Key metrics for the LLM model include an MSE of 2.0538, RMSE of 1.4331, MAE of 0.7142, SD of 1.4287, Correlation Coefficient (R) of 0.4197, and Coefficient of Determination (R^2) of 0.0244. While the LLM approach exhibits better contextual understanding, the results highlight that its performance is comparable to the SVM model. The findings indicate that while LLMs offer the advantage of leveraging advanced contextual embeddings, traditional machine learning approaches like SVM remain viable, lightweight alternatives for structured email classification. The comparative analysis underscores the strengths and limitations of both approaches, offering insights into their practical applications and deployment scenarios.

Keywords

Email Classification, Machine Learning, Support Vector Machine, Large Language Model, Natural Language Processing.