Sentiment Analysis of Customer Satisfaction Across 48 U.S. Banks

Ruoyi Li
Department of Math
University of Michigan
Ann Arbor, USA
ruoyil@umich.edu

***Abstract*—Sentiment analysis has emerged as a vital tool in the financial sector, and the Long Short-Term Memory (LSTM) model plays a crucial role in it. This study aims to use LSTM to analyse public sentiment derived from the customer reviews. Additionally, this research evaluates the predictive performance and practical impact of LSTM-based sentiment analysis, and highlights its potential in forecasting market behavior.**

***Index Terms*—Sentiment Analysis, LSTM, Imbalanced Data, Text Classification, Cross-Entrop Loss in LSTM, Stratified Sampling for Splitting**

I. Introduction

With banks managing millions of customer interactions, a sentiment analysis enables institutions to detect early signs of dissatisfaction, reduce customer churn, and enhance loyalty. Based on one study from ResearchGate, it indicates approximately 80% of the world's data is unstructured and not organized during a pre-defined manner, and most of them come from text data, like reviews [1]. Additionally, the LSTM network has been widely explored for analyzing market trends and beyond, because their ability to handle sequential data[2].

The motivation behind this project stems from a strong interest in exploring the widespread application of the LSTM model in sentiment analysis. This project has two primary objectives. The first is to evaluate the performance of the LSTM model, with a focus on its ability to accurately classify sentiment in customer reviews. The second is to learn the relationship between top banks and customer satisfactions.

II. METHOD

A. Dataset Description, Selection and Cleaning

The dataset is a collection of 20,000 customer reviews on 48 U.S. banks. The raw data contains 7 variables, which are `author`, `date`, `location`, `bank`, `star`, `text`, and `like`. However, in this project, only three elements are used: `bank` (Consist of 48 different U.S. banks); `star` (1 and 2 represent poor performance, 3 is neutral, and 4 and 5 are good performance); `text` (positive, neutral, or negative comments).

To ensure the dataset is ready for both text classification (LSTM) and statistical analysis, this project handles missing values by removing incomplete reviews or star ratings. Then, star ratings will be mapped to the three sentiment classes:

TABLE I
STAR RATING TO SENTIMENT MAPPING

Star	Sentiment
1–2	Negative (0) – Poor Performance
3	Neutral (1) – Neutral Performance
4–5	Positive (2) – Good Performance

B. Problem Formulation

Analyzing thousands of reviews manually is impractical, so this project aims to automate identifying the classifications of customer sentiments based on their written reviews by using the LSTM model and text classification technique. This sentiment analysis is a supervised multi-class text classification problem, and to build a pipeline, the followings are needed:

- **Training set and Testing set**: The dataset is divided into a training set (80% of the dataset) and the testing set (20% of the dataset)
- **Input**: The inputs are `text`, and `star`. The model learns to classify sentiment as positive, neutral or negative by being trained using labels derived from `star`.
- **Output**: The first output is the sentiment class (i.e. 0, 1, 2) and the other refers to financial metrics outputs (i.e. positive ratio and negative ratio).

C. Model Formulation

In this project, an LSTM-based model to perform text classification is used. In addition, 80% of the dataset are the training set, and the rest 20% are the testing set.

Manipulations are performed on the text reviews by converting them into structured numerical sequences. Each review text x_i is first preprocessed and tokenized into a sequence of tokens. Then, each token is transformed into a dense embedding vector using an embedding layer. Moreover, the embedded sequence is passed to a single-layer LSTM network which can identify the dependencies among each tokens, and then the LSTM-derived vector is passed to a linear layer to produce an output corresponding to the three sentiment classes. During the LSTM model learning process, the following cross-entropy loss function plays a central role:

$$L = -\sum_{i=1}^{N} y_i \cdot \log(\hat{y}_i)$$

where y_i is the true sentiment label (derived from star ratings), and $\hat{y}_i$ is the predicted probability distribution produced by the model for class i.

The key idea is that the closer the predicted probability is to the correct label, the smaller the loss is. Besides, there are three major methodologies throughout the project. In the data pre-processing and labeling steps, tokenization, embedding, and padding are used. In terms of the modeling process, the LSTM-based model is selected to perform text classification. In the training step, the Cross-Entropy loss function is used to minimize the loss and evaluate the trained model.

III. Result

A. Model Setup

The model architecture and parameters were tuned based on the performance on a bank review dataset. The following setups are used:

- A LSTM layer with a hidden dimension of 128 units, followed by a fully connected output layer.
- A maximum sequence length of 100 tokens
- The model is trained for 30 epochs
- The dataset was split into 80% training and 20% testing

B. Performance Evaluation and Comparison Analysis

The Fig.1 derived from the loss function indicates the model is trained well, as the loss values are reaching to zero.

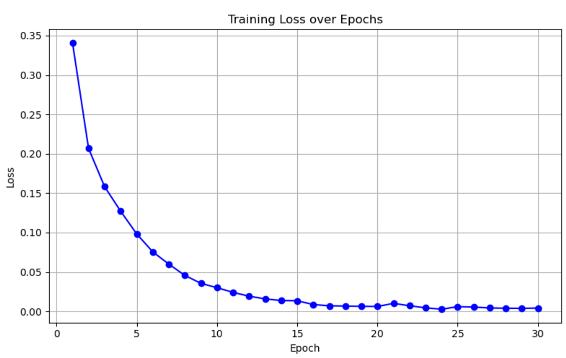

Fig. 1. Loss Function Result

The Fig.2 indicates, the LSTM model performs strongly on classifying negative reviews, achieving high accuracy with 3308 correct classifications. However, it struggles with neutral sentiment, likely due to its subtle linguistic cues. The model performs moderately well on positive reviews. This pattern may be influenced by class imbalance, where negative reviews dominate the dataset. Future work could benefit from techniques like resampling, class weighting, or using pre-trained transformer models to better capture nuanced sentiments.

By analysing the below two tables, there is moderate agreement between sentiment analysis and star rating trends, especially for banks with more reviews. For example, Merrick Bank emerges as a high-confidence performer based on both data types. When interpreting results from banks with very few reviews, a single perfect score can inflate rankings, which

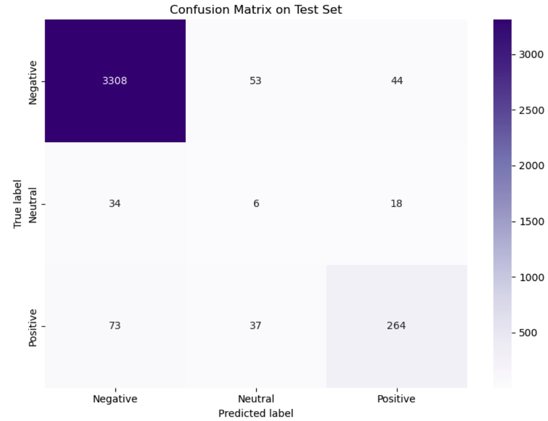

Fig. 2. Confusion Matrix Result on Test Set

might cause a misleading conclusion. Therefore, a larger and more comprehensive dataset should be used to create a better model performance.

TABLE II
Top 5 Banks by Positive Sentiment Ratio

Bank	Positive	Neutral	Negative	Positive Ratio
bmo-harris-bank	1	0	0	1.000000
merrick_bank	141	15	74	0.613043
arrowhead_credit_union_ca	2	0	2	0.500000
armed_forces_bank	1	1	5	0.142857
usaa_banking	11	4	71	0.127907

TABLE III
Top 5 Banks by Positive Star Rating Ratio

Bank	Good Reviews	Neutral Reviews	Poor Reviews	Positive Ratio
bmo-harris-bank	1	0	0	1.000000
arrowhead_credit_union_ca	3	0	1	0.750000
merrick_bank	145	16	69	0.630435
comerica_bank	5	0	18	0.217391
sofi-money	3	0	15	0.166667

IV. Conclusion

Based on the analysis of both sentiment-based and star-rating-based feedback, big banks like Merrick Bank provides a more accurate results when examining the model performance, while smaller banks, like BMO Harris Bank have low volumes of reviews, limiting the reliability of that result. Banks like USAA Banking and Sofi Money reveal discrepancies between sentiment and rating, indicating potential inconsistencies in customer expectations or how they express dissatisfaction. In addition, it is obvious that the top bank Merrick Bank has far more reviews and ratings, so the relatively accurate outcomes indicates it is more likely there is a positive relationship between customer satisfactions and institution reputations.

References

[1] W. Liu, J. Li, Z. Li, and Z. Sun, "Text-based Sentiment Analysis using LSTM," *2020 International Conference on Computer Science, Engineering and Education Applications (ICCSEEA)*, Kyiv, Ukraine, 2020, pp. 293–299.

[2] L. Bacco, L. Petrosino, D. Arganese, and L. Vollero, "Investigating Stock Prediction Using LSTM Networks and Sentiment Analysis of Tweets Under High Uncertainty," *IEEE Access*, vol. PP, pp. 1–1, Jan. 2024.