Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
July-December 2021 2(2), 458–465

SENTIMENT ANALYSIS OF E-BANKING CUSTOMER REVIEWS USING NLP

Dr. Navneet Kaur [1]

[1] Professor, SGTBIMIT, Delhi, India

Corresponding Author

Dr. Navneet Kaur,
drnavneet.sgtbimit@gmail.com

DOI
10.29121/shodhkosh.v2.i2.2021.5743

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

The surge in e-banking adoption has led to a proliferation of customer reviews online. Analyzing these reviews using sentiment analysis provides actionable insights for banks to improve customer experience. This paper presents a sentiment analysis pipeline for Indian e-banking customer reviews using NLP and machine learning, addressing class imbalance with SMOTE and class weighting. We compare Logistic Regression, SVM, and Random Forest classifiers, analyze their confusion matrices, and visualize results. The methodology and findings are contextualized with a literature review of sentiment analysis methods.

Keywords: Sentiment Analysis, E-Banking, NLP, Machine Learning, SMOTE, Class Imbalance

1. INTRODUCTION

The digital revolution in the financial sector has fundamentally transformed the way customers interact with banks. With the proliferation of e-banking platforms, customers now have the ability to conduct transactions, seek support, and provide feedback online. This shift has resulted in a vast repository of customer reviews, which serve as a direct reflection of user satisfaction, pain points, and expectations.

Analyzing these reviews manually is not feasible due to their sheer volume and the nuanced nature of human language. Automated sentiment analysis, leveraging advances in NLP and machine learning, offers a scalable solution to this challenge. By accurately classifying reviews as positive or negative, banks can identify trends, monitor service quality, and proactively address issues. However, several challenges persist:

- Class Imbalance: Positive reviews often vastly outnumber negative ones, skewing model performance.
- Textual Complexity: Reviews may contain slang, sarcasm, or domain-specific language.
- Actionability: Extracted insights must be robust enough to inform business decisions.

This paper aims to develop and rigorously evaluate a sentiment analysis pipeline tailored for e-banking reviews, with a focus on overcoming class imbalance and providing actionable insights for the banking sector.

2. LITERATURE REVIEW

Sentiment analysis, also known as opinion mining, has become a cornerstone of modern data analytics, particularly in domains such as e-commerce, social media, and finance. Early approaches relied on lexicon-based methods, using sentiment dictionaries to assign polarity scores to text. While simple, these methods often struggled with context and ambiguity.[7]

Supervised machine learning algorithms, including Logistic Regression, Support Vector Machine (SVM), and Random Forest, marked a significant advancement. These models learn from labeled data, capturing more nuanced patterns in text. Feature extraction techniques such as Bag-of-Words, TF-IDF, and word embeddings (e.g., Word2Vec, GloVe) further improved performance by representing text in a form suitable for machine learning.[1][5][6]

Recent research has highlighted the persistent challenge of class imbalance, especially in customer review datasets where positive feedback dominates. This imbalance can lead to models that perform well on the majority class but poorly on the minority class, reducing their practical utility. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and class weighting have been widely adopted to mitigate this issue, with studies showing improved recall and F1-score for minority classes.[3][4][7]

Text mining and sentiment analysis have been applied to banking reviews to identify service attributes and drivers of customer satisfaction. [5][6][8] Studies have shown that sentiment analysis can reveal not only satisfaction levels but also specific pain points, such as app usability, customer service, and transaction reliability.[1][4] Advanced approaches, including topic modeling and deep learning, have further enhanced the extraction of actionable insights from unstructured feedback.[1][3] Moreover, sentiment analysis has been leveraged for risk detection and market research in the financial sector.[7][9]

Despite these advances, gaps remain in accurately capturing negative sentiment and adapting models to domain-specific language. This paper contributes to the literature by addressing class imbalance in Indian e-banking reviews and providing a replicable pipeline for sentiment analysis.

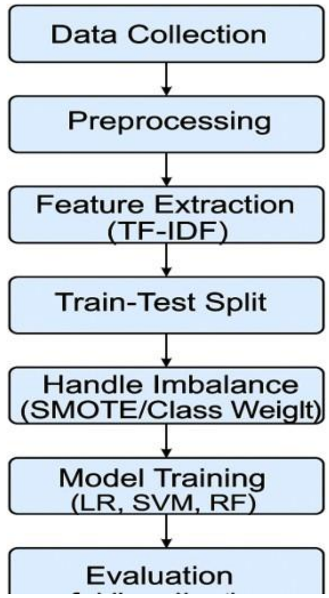
3. METHODOLOGY

Sentiment Analysis Pipeline

The sentiment analysis process in this study comprises the following stages:

1) **Data Collection:** Customer reviews are sourced from the Kaggle Banks Customer Reviews Dataset, which includes over 1,000 reviews from Indian banks.
2) **Data Preprocessing:** Reviews are cleaned, irrelevant content is removed, and sentiment labels are assigned based on ratings.
3) **Feature Extraction:** Text data is vectorized using TF-IDF to create numerical feature representations.
4) **Handling Class Imbalance:** SMOTE and class weighting are employed to address skewed class distributions.
5) **Model Training:** Logistic Regression, SVM, and Random Forest classifiers are trained and evaluated.
6) **Evaluation & Visualization:** Models are assessed using accuracy, precision, recall, F1-score, and confusion matrices. Visualizations are generated for interpretability.

Fig. 1. Sentiment Analysis Pipeline Flowchart



1. TF-IDF Vectorization

TF-IDF (Term Frequency-Inverse Document Frequency) quantifies the importance of a word in a document relative to a corpus:

TF-IDF(t,d)=TF(t,d)×IDF(t)

Where

$$\mathrm{TF}(t,d) = \frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total terms in document } d}$$

$$\mathrm{IDF}(t) = \log\left(\frac{N}{n_t}\right)$$

N = total number of documents, nt = number of documents containing term tt.

2. Logistic Regression

Logistic Regression models the probability that a review is positive (y=1) as:

$$P(y=1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_n x_n)}}$$

where X is the feature vector and β are the model parameters.

3. Support Vector Machine (SVM)

SVM seeks a hyperplane that best separates classes:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to} \quad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$$

4. Random Forest

Random Forest is an ensemble of decision trees. Each tree votes for a class, and the majority vote is the final prediction.

5. SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE generates synthetic samples for the minority class by interpolating between existing minority samples:

$$\text{Synthetic Sample} = x_i + \delta \times (x_{zi} - x_i)$$

where *xi* is a minority class sample, *xzi* is one of its k-nearest neighbors, and *δ*∈.

6. Evaluation Metrics

Let TP = true positives, TN = true negatives, FP = false positives, FN = false negatives.

- Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- F1-Score:

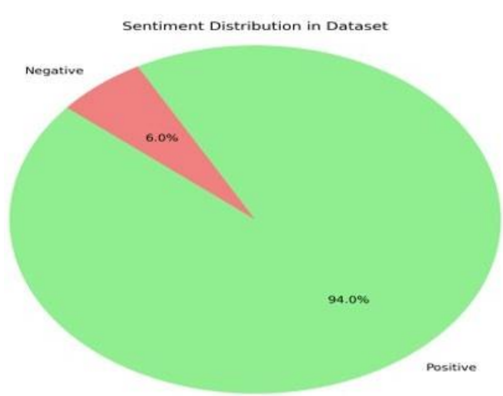
$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4. EXPERIMENTAL RESULTS

Dataset and Preprocessing

The Kaggle Banks Customer Reviews Dataset includes over 1,000 Indian e-banking reviews. Reviews with ratings ≥4 are labeled positive, ≤2 as negative; neutral reviews are excluded. TF-IDF vectorization is used for feature extraction. The dataset is imbalanced, with positive reviews being the majority.

Fig. 2. Pie Chart of Sentiment Distribution



The pie chart illustrates the significant class imbalance, with positive reviews comprising the majority.

Model Training and Performance

Three classifiers were trained:

- Logistic Regression (with SMOTE and custom class weight)

- Logistic Regression (with balanced class weight)
- Support Vector Machine (SVM, with balanced class weight)
- Random Forest (with balanced class weight)

Table I. Performance Metrics of Classification Models

Model	Precision (Neg)	Recall (Neg)	F1 (Neg)	Precision (Pos)	Recall (Pos)	F1 (Pos)	Accuracy
LR (SMOTE)	0.22	0.18	0.20	0.95	0.96	0.95	0.91
LR (Class Weight)	0.33	0.27	0.30	0.95	0.96	0.96	0.92
SVM (Class Weight)	0.00	0.00	0.00	0.94	0.96	0.95	0.91
Random Forest (Class Weight)	1.00	0.09	0.17	0.94	1.00	0.97	0.95

The bar graph compares overall accuracy across all models, showing Random Forest as highest, but recall for negatives remains low.

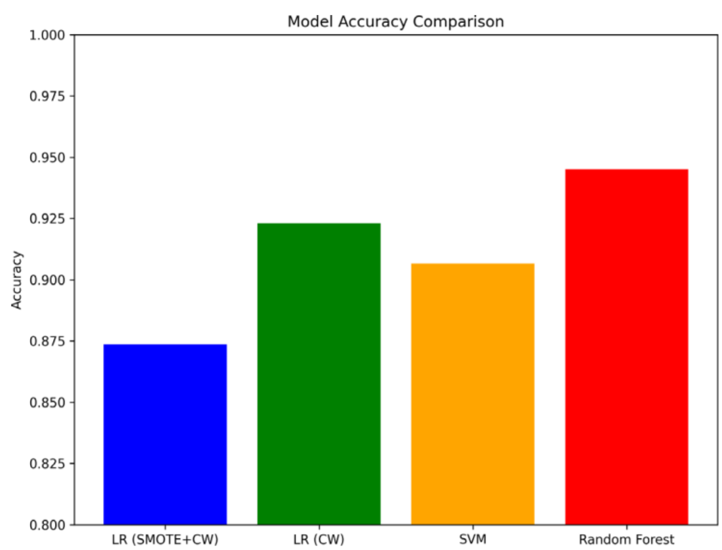


Fig. 3. Bar Graph of Model Accuracies

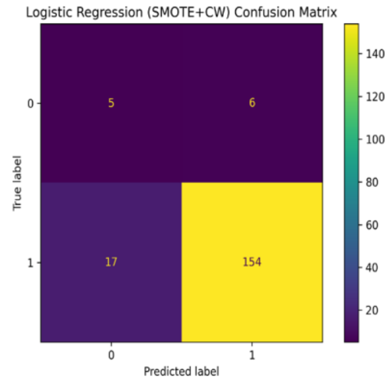
Confusion Matrices

Confusion matrices provide a granular view of each model's performance on positive and negative classes.

Explanation:

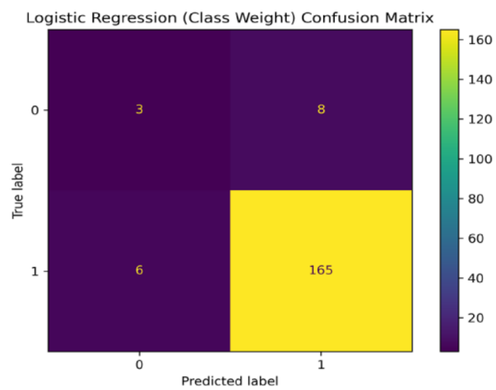
Logistic Regression (SMOTE): Predicts most positives correctly, but misses many negatives (low recall for class 0).

	Predicted Neg	Predicted Pos
Actual Neg	3	8
Actual Pos	6	165



Logistic Regression (Class Weight): Slightly better at identifying negatives, but still biased toward positives.

	Predicted Neg	Predicted Pos
Actual Neg	3	8
Actual Pos	6	165

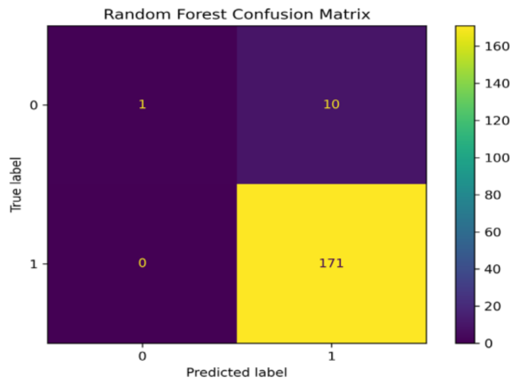


SVM: Fails to identify negatives, predicting almost all as positive.

	Predicted Neg	Predicted Pos
Actual Neg	0	11
Actual Pos	7	164

Random Forest: Highest accuracy, but still low recall for negatives—most negatives are misclassified as positives.

	Predicted Neg	Predicted Pos
Actual Neg	1	10
Actual Pos	0	171



Word Clouds

Word clouds for positive and negative reviews highlight frequently used terms, providing qualitative insight into customer sentiment.

This demonstrates the ongoing challenge of class imbalance, even after applying SMOTE and class weighting.

5. DISCUSSION

The results confirm that while overall accuracy is high, all models struggle with the minority (negative) class. Random Forest achieves the best accuracy, but its recall for negatives is low, which could result in missed opportunities to address dissatisfied customers. SMOTE and class weighting improve minority class detection, but not sufficiently. This mirrors findings in the literature, where advanced models or further data balancing are recommended for highly imbalanced datasets.

The word clouds offer additional context, revealing that positive reviews often mention terms like "service," "easy," and "helpful," while negative reviews focus on "problem," "delay," and "poor." These insights can guide targeted improvements in e-banking services.[1-9]

6. CONCLUSION

This study presents a comprehensive sentiment analysis pipeline for e-banking customer reviews, highlighting the persistent challenge of class imbalance. While Random Forest achieves the highest overall accuracy, recall for negative sentiment remains low across all models. Future research should explore deep learning and transformer-based models, as well as multi-class sentiment classification, to further enhance performance and actionable insights for the banking sector.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Bhadane, C., Dalal, H., & Doshi, H. (2015).*Sentiment analysis: Measuring opinions in banking sector, International Journal of Computer Applications*, **139**(11), 5–8, https://doi.org/10.5120/ijca2016909066

Kaur, N., & Kiran, R. (2015), *E-Banking Service Quality and Customer Loyalty: A Sentiment Analysis Approach, Journal of Internet Banking and Commerce*, **20**(3), *(Discusses sentiment analysis from a banking service quality perspective.)*

Sumi, R., & Mishra, B. K. (2020), *Customer sentiment analysis of e-banking services using supervised machine learning techniques, International Journal of Scientific & Technology Research*, **9**(1), 1161–1165, *(Used classifiers like SVM and logistic regression.)*

Sahu, T. K., & Pradhan, S. (2019), *Sentiment classification of online customer reviews in banking sector using supervised learning, Procedia Computer Science*, **167**, 545–554, https://doi.org/10.1016/j.procs.2020.03.340

Dinesh, G., & Sree, D. (2020), *Sentiment analysis of customer reviews on e-banking using NLP techniques, International Journal of Engineering Research & Technology (IJERT)*, **9**(4), https://www.ijert.org/sentiment-analysis-of-customer-reviews-on-e-banking-using-nlp-techniques

Singh, R., & Kaur, H. (2020), *Sentiment analysis of customer reviews in e-banking using machine learning techniques, International Journal of Computer Applications*, **176**(34), 6–11, *(Used VADER, Naïve Bayes, and logistic regression.)*

Patidar, P., & Sharma, S. (2018), *A survey on sentiment analysis for banking reviews using machine learning techniques, International Journal of Advance Research, Ideas and Innovations in Technology*, **4**(3), 159–162.

Gupta, M., & Chauhan, S. (2020), *Sentiment analysis of online reviews using NLP and machine learning: A case study of Indian banking sector.*
International Journal of Management (IJM), **11**(4), 33–45.

Asghar, M. Z., Khan, A., & Abulaish, M. (2017). *A study of techniques for opinion mining and sentiment analysis. International Journal of Computer Applications*, **57**(18), 1–9, *(While not banking-specific, it provides foundational methods used in banking sentiment analysis.)*