Journal Name

Research Paper

Analyzing Public Sentiment on Demonetization Using SVM: A Machine Learning Approach

Kaliappan M. [1], **Guruprakash B.** [2], **Rajalakshmi**[3], **J. Blessing Karunya T.** [4], **Mariappan E.** [1], **Ramnath M.** [1], **Angel Hepzibah R.** [1]

[1]*Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India*
[2]*Department of Computer Science and Engineering (AI&ML), Sethu Institute of Technology, Virudhunagar, Tamil Nadu, India*
[3]*Department of Electronics and Communication Engineering, Sethu Institute of Technology, Virudhunagar, Tamil Nadu, India*
[4]*Department of Information Technology, P.S.R. Engineering College, Sivakasi, Virudhunagar, Tamil Nadu, India*

Article history
Received: 6 December 2024
Revised: 5 April 2025
Accepted: 24 April 2025

*Corresponding Author:
Dr. Mariappan E
Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, India
Email: mapcse.e@gmail.com

Abstract: The Indian economy experienced significant disruption following the implementation of demonetization, a policy initiative aimed at eliminating black money, controlling inflation, and promoting financial inclusion. However, this currency ban generated widespread debate and polarized public opinion. This study analyzes public sentiment toward demonetization using social media data, specifically Twitter posts characterized by mixed sentiments, sarcasm, and nuanced linguistic expressions. We employ a PAD-SVM (Preprocessing-Analysis-Decision Support Vector Machine) approach comprising three stages: preprocessing, descriptive analysis, and prescriptive analysis. The preprocessing stage involves data cleaning, handling missing values, and feature extraction from tweet data. The descriptive analysis stage identifies key influencers and performs exploratory data analysis related to demonetization discourse. Subsequently, sentiment analysis is conducted to quantify user sentiments and assign polarity scores to individual tweets. Predictive modeling is then applied to forecast evolving public perception toward demonetization over time. This approach combines machine learning, statistical modeling, and natural language processing (NLP) techniques to process unstructured textual data and classify sentiments as positive, negative, or neutral. The integration of sentiment analysis with predictive analytics provides valuable real-time insights into public opinion dynamics and enables future trend forecasting regarding major economic policy interventions.

Keywords: Demonetization, Sentiment Analysis, Support Vector Machine, Predictive Analytics, Natural Language Processing, Social Media Analytics, Twitter Data Analysis

Introduction

Demonetization refers to the process of removing a currency's status as a legal medium of exchange. Governments implement such measures to counter corruption, tax evasion, counterfeit currency, and to promote digital economies. In November 2016, India's government demonetized its highest denomination currency notes (₹500 and ₹1000) creating immediate economic and societal impacts. As 78% of transactions were cash-based, the public response was intense and widely voiced through platforms like Twitter.

Analyzing public sentiment during such events is essential for gauging the socio-economic impact and policy effectiveness. Predictive analytics, powered by AI, ML, and data mining, enables the extraction and analysis of sentiment from such large-scale unstructured data. This review focuses on techniques and frameworks used for sentiment classification and predictive modeling using Twitter data during events like demonetization.

Literature Review

Various researchers have investigated the use of predictive analytics and sentiment analysis for extracting insights from Twitter data. Researchers proposed an intelligent framework that integrates structured and unstructured data sources, including Twitter and YouTube, to predict box office revenues. Their system demonstrated that even users without domain expertise could effectively utilize the platform for accurate predictions.

In the educational context, Gulati (2015) applied predictive modeling using decision trees to identify the factors contributing to student dropout rates. Her work emphasized the importance of preprocessing; particularly feature selection, to enhance the accuracy and efficiency of classification models.

Lin & Kolcz (2012) developed a scalable sentiment analysis system using Hadoop to process over a million tweets annotated with emoticons. Their model achieved accuracy between 77% and 82%, showcasing the effectiveness of distributed systems in handling real-time social media data through Java-Pig integration.

Bian et al. (2012) explored Twitter as a tool for surveillance of adverse drug events. Using a massive dataset comprising over 2 billion tweets, they implemented a Support Vector Machine (SVM) classifier that achieved an AUC of 0.82 and accuracy of 0.74, demonstrating the potential of predictive analytics for public health applications.

Liu et al. (2013) proposed a sentiment analysis system utilizing the Naive Bayes classifier implemented in a Hadoop environment. Their model processed movie review data and achieved an accuracy of 80.85%, suggesting that even basic probabilistic models can perform well in big data contexts.

Cuesta et al. (2014) introduced an extensible, open-source framework for Twitter data analysis using MongoDB and Python. Their system supported three-class sentiment categorization and offered scalability for large-scale social media sentiment analysis workflows.

In the financial domain, Skuza & Romanowski (2015) leveraged sentiment analysis to forecast stock price movements. They employed a Map Reduce-based Naive Bayes classifier to process tweets associated with specific companies over a three-month period. Their findings revealed a significant correlation between the sentiments expressed in tweets and subsequent stock performance.

Similarly, Tare et al. (2014) developed a tweet classification pipeline using a Hadoop-based Map Reduce framework integrated with the Twitter4j API. Their approach effectively managed large volumes of tweet data, showcasing the advantages of distributed computing for scalable and efficient social media sentiment classification.

Collectively, these works highlight the effectiveness of predictive models and distributed systems for extracting actionable insights from Twitter data. They provide a strong foundation for developing sentiment analysis tools aimed at understanding public opinion during socio-economic events such as demonetization. Vimal et al. (2021) employed a Random Forest algorithm integrated with Artificial Intelligence techniques to forecast influenza outbreak patterns using data extracted from Twitter. Their approach demonstrated the potential of social media analytics in early disease detection by effectively identifying flu trends based on user-generated content.

Vimal et al. (2021) proposed an intelligent medical prediction framework for glaucoma progression detection by leveraging the K-means clustering algorithm in combination with the Gray-Level Co-occurrence Matrix (GLCM) technique. This hybrid approach effectively enhances the accuracy of glaucoma diagnosis by extracting and analyzing key texture features from retinal images, contributing to advancements in smart healthcare systems.

Proposed Work

SVM Algorithm for Sentiment Analysis

Data Preparation

Demonetization tweets are collected and pre-processed 1. Cleaning and 2. removing stop words. Each tweet is represented as a feature vector using TF-IDF). Tweets are labelled with sentiment such as positive, negative, or neutral.

Finding the Hyperplane

SVM aims to find the best hyperplane that separates data points of different sentiments with the largest margin. The hyperplane is defined by a weight vector (**w**) and a bias (**b**). The equation of the hyper plane is:

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0$$

where x is the feature vector of a tweet.

Maximizing the Margin

The margin is the distance between the hyperplane and the nearest data points called support vectors. The objective of SVM is to find the optimal hyperplane that provides the largest separation between data classes, ensuring better generalization and stability.

Handling Non-Linearly Separable Data

SVM handles non-linearly separable data by applying a kernel function that maps input features into a higher-dimensional space for easier classification. These schemes determine the optimal cluster head by evaluating both distance and energy parameters. The Enhanced Immune Genetic Algorithm (EIGA) is employed to preserve population diversity, while the Memory-Enhanced Genetic Algorithm (MEGA) is utilized to retain information about previous environmental states (Michal & Romanowski, 2015; Wook et al., 2009) Common kernels include linear, polynomial, and radial basis function (RBF). Early detection or screening of changes in retinal images is particularly challenging and time-consuming for ophthalmologists, as the initial variations in size and color closely resemble the surrounding blood vessels in the retina (Taboada, 2016).

It improves the client's utilization of cloud data storage in a secure manner, while also improving global data storage efficiency through heuristic techniques. The K-Nearest Neighbors (KNN) algorithm is employed to enable efficient query searching within data clusters by the client (Tushar & Srivastava, 2012). The server provides a distributed public key using RSA to ensure security and address memory recycling issues. This approach improves the client's secure access to cloud data storage and enhances global storage efficiency through the implementation of heuristic techniques (Wilson et al, 2005).

Classification

Once the hyperplane is found, new tweets are classified based on which side of the hyperplane they fall on.

If $\mathbf{w} \cdot \mathbf{x} + \mathbf{b} > 0$, the tweet is classified as positive.

If $\mathbf{w} \cdot \mathbf{x} + \mathbf{b} < 0$, the tweet is classified as negative.

If $\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0$, the tweet is on the hyperplane and can be classified as neutral.

Mathematical Model

The optimization problem for SVM can be formulated as:

```
Minimize: 1/2 ||w||^2
Subject to: yi(w · xi + b) >= 1 for all i
```

||w||^2: Represents the squared norm of the weight vector, which is minimized to maximize the margin.

yi: Represents the sentiment label of the i-th tweet (+1 for positive, -1 for negative).

xi: Represents the feature vector of the i-th tweet.

b: Represents the bias term.

In data preparation, stop words were removed in demonetization tweets data. Fig. 1 shows that neutral tweets are 8803, negative tweets are 3834 and positive tweets are 3207.

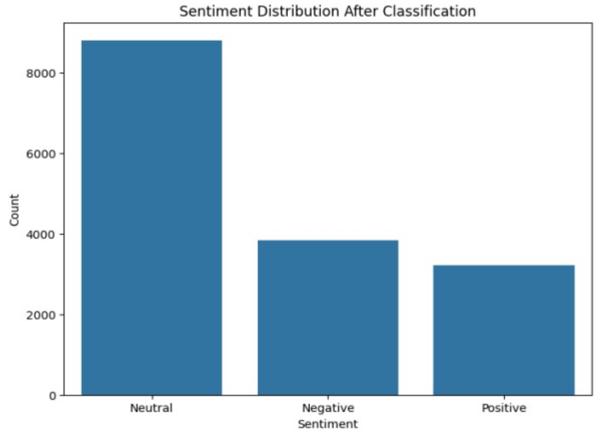

Fig. 1. Demonetization tweets data

Confusion Matrix of Predicted value

Table 1 indicates that 118 negative tweets were correctly classified, while 15 were misclassified as neutral and 7 as positive. Figure 2 presents the confusion matrix for the PAD-SVM model.

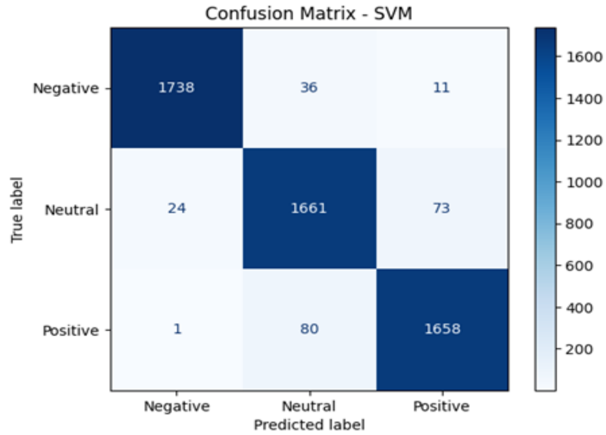

Fig. 2. Confusion Matrix of PAD-SVM

Table 1. Confusion Matrix

Predicted	Negative	Neutral	Positive
Negative	80	15	6
Neutral	9	142	19
Positive	6	13	111

Table 2. Confusion Matrix

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.98	0.97	0.98	1785
Neutral	0.95	0.93	0.94	1758
Positive	0.94	0.97	0.95	1739
Accuracy	0.96	0.96	0.96	5282
Macro Avg.	0.96	0.96	0.96	5282
Weighted Avg.	0.96	0.96	0.96	5282

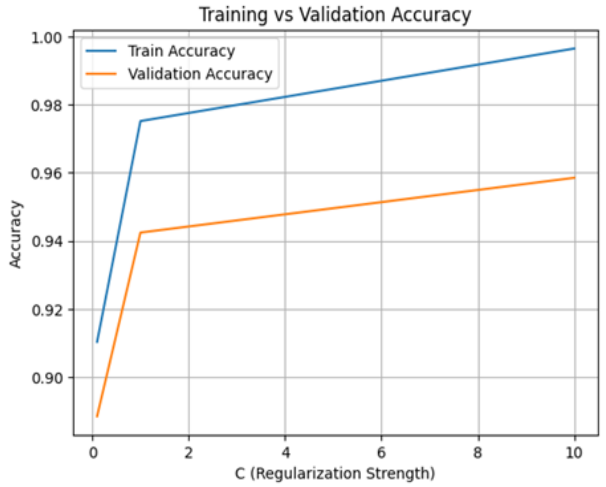

Fig. 3. Training vs Validation Accuracy

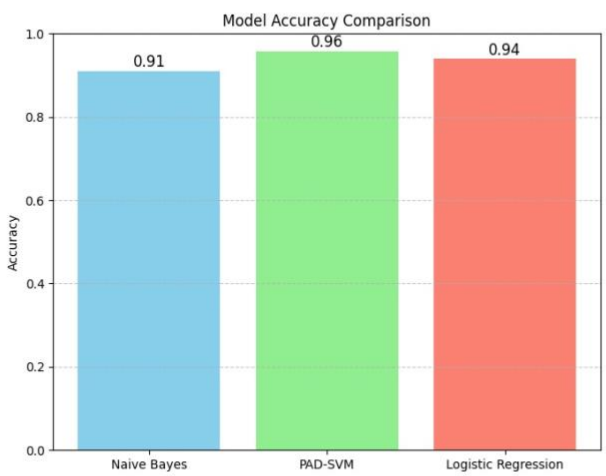

Fig. 4. Accuracy competitions of Naïve Bayes, PAD-SVM and Logistic Regression

SVM is handling Complexity such as Kernels and Margin Maximization: SVM's flexibility with kernels (e.g., RBF) allows it to model non-linear relationships between words and sentiments, capturing the nuances in demonetization tweets. This adaptability addresses the challenge of complex sentiment expressions. SVM focuses on finding the hyperplane with the largest margin, leading to better generalization and robustness. This principle helps reduce over fitting, particularly in the presence of complex and potentially noisy data. RBF kernel is often preferred for text data as it can capture non-linear relationships between words and sentiments

SVM handle well High-Dimensional Data. SVM incorporates regularization controlled by the C parameter to prevent over fitting in high-dimensional spaces. This parameter helps balance model complexity and generalization performance, ensuring effective handling of the large feature space inherent in text data. The regularization parameter plays a critical role in controlling the balance between margin maximization and classification error during model training. It needs to be tuned based on the dataset

Also, it Addressing Class Imbalance by assign different weights to classes during training using the class weight parameter, addressing the issue of imbalanced data. This helps the model give more importance to under-represented classes like neutral sentiment, potentially improving overall accuracy. If class imbalance is significant, assigning appropriate weights to classes can improve performance

Consider the following demonetization tweets:

- **Positive:** "Demonetization is a bold step towards a corruption-free India."
- **Negative:** "Demonetization has caused immense hardship to common people."
- **Neutral:** "I'm waiting to see the long-term effects of demonetization."

SVM with an RBF kernel can capture the complex relationships between words and sentiments in these tweets. The C parameter can be tuned to balance model complexity and generalization, while class weights can address potential imbalance between sentiment categories. SVM, with its ability to handle complex sentiment expressions, high-dimensional data, and class imbalance, is a strong candidate for sentiment analysis on demonetization data. By carefully tuning parameters like kernel, C, and class weights, we can achieve higher accuracy and gain valuable insights into public opinion on this crucial policy.

Conclusion

This study analyzed public sentiment regarding India's 2016 demonetization policy using Twitter data to classify posts into three sentiment categories: Positive, Negative, and Neutral. Following comprehensive data preprocessing and TF-IDF vectorization, multiple machine learning classifiers were trained and evaluated. The Support Vector Machine (SVM) classifier demonstrated superior performance, achieving 96% accuracy, thereby establishing itself as the most reliable model for sentiment classification in this context.

Performance evaluation using a confusion matrix revealed robust classification capabilities across all sentiment categories. The model correctly identified 118 negative tweets, 142 neutral tweets, and 111 positive tweets, with minimal misclassifications. While some confusion existed between closely related sentiment categories, particularly Neutral and Positive, the heatmap visualization provided valuable insights into the model's classification patterns and areas for potential refinement. These findings demonstrate the effectiveness of SVM-based sentiment analysis for understanding public opinion on major economic policy interventions and establish a methodological framework applicable to similar social media analytics studies.

Ethics

This study represents the authors' original research and has not been previously published or submitted for publication elsewhere. All research was conducted in accordance with ethical standards for data collection and analysis of publicly available social media data.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare no competing financial or non-financial interests that could inappropriately influence this work.

Funding Information

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

Acknowledgements

The authors express their gratitude to the anonymous reviewers for their constructive feedback and insightful comments that improved the quality of this manuscript.

Author Contributions

Kaliappan M: Conceptualization, Methodology, Investigation, Experimentation, Writing - Original Draft
Guruprakash B: Writing - Review and Editing
Rajalakshmi: Data Collection, Formal Analysis
J. Blessing Karunya T: Validation, Writing - Review and Editing
Mariappan E: Data Curation, Formal Analysis, Supervision, Project Administration
Ramnath M: Writing - Original Draft, Visualization
Angel Hepzibah R: Writing - Review and Editing

References

Bian, J., Topaloglu, U., & Yu, F. (2012). Towards large-scale Twitter mining for drug-related adverse events. Proceedings of the 2012 International Workshop on Smart Health and Wellbeing (SHB '12), 25-32.

Cuesta, A., David, F., & Rodríguez-Moreno, M. D. (2014). A framework for massive Twitter data extraction and analysis. *Malaysian Journal of Computer Science, 3*, 50-67.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM, 51*(1), 107-113. https://doi.org/10.1145/1327452.1327492

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266-6282. https://doi.org/10.1016/j.eswa.2013.05.057

Gulati, H. (2015). Predictive analytics using data mining technique. 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), 713-716.

Hao, M., Rohrdantz, C., Janetzko, H., Dayal, U., Kiem, D. A., Haug, L. E., & Hsu, M. C. (2014). Visual sentiment analysis on Twitter data streams. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology* (pp. 277-286). IEEE.

Jose, A. K., Bhatia, N., & Krishna, S. (2010). Twitter sentiment analysis. *IEEE*, 1-5.

Kaliappan, M., Mariappan, E., Prakash, M. V., & Paramasivan, B. (2016). Load balanced clustering technique in MANET using genetic algorithms. *Defence Science Journal, 66*(3), 251-258.

Lin, J., & Kolcz, A. (2012). Large-scale machine learning at Twitter. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 793-804). https://doi.org/10.1145/2213836.2213958

Liu, B., Blasch, E., Chen, Y., Shen, D., & Chen, G. (2013). Scalable sentiment classification for big data analysis using Naive Bayes classifier. In *2013 IEEE International Conference on Big Data* (pp. 99-104). https://doi.org/10.1109/BigData.2013.6691679

Márquez-Vera, C., Morales, C. R., & Soto, S. V. (2013). Predicting school failure and dropout by using data mining techniques. *IEEE Journal of Latin-American Learning Technologies, 8*(1), 7-14.

Michal, S., & Romanowski, A. (2015). Sentiment analysis of Twitter data within big data distributed environment for stock prediction. In *Proceedings of the Federated Conference on Computer Science and Information Systems* (Vol. 2, pp. 1349-1354). https://doi.org/10.15439/2015F241

Mohit, T., Gohokar, I., Sable, J., Paratwar, D., & Wajgi, R. (2014). Multi-class tweet categorization using MapReduce paradigm. *International Journal of Computer Trends and Technology, 3*, 78-81.

Rohrdantz, C., Hao, M., Dayal, U., Haug, L. E., & Hsu, M. C. (2010). HaLoop: Efficient iterative data processing on large clusters. *Proceedings of the VLDB Endowment, 3*(1-2), 285-296. https://doi.org/10.14778/1920841.1920883

Sivaram, M., Kaliappan, M., Shobana, S. J., Prakash, M. V., & Porkodi, V. (2023). Secure storage allocation scheme using fuzzy-based heuristic algorithm for cloud. *Journal of Ambient Intelligence and Humanized Computing, 14*, 1-9.

Skuza, M., & Romanowski, A. (2015). Sentiment analysis of Twitter data within big data distributed environment for stock prediction. In Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on, pp. 1349-1354. IEEE.

Taboada, M. (2016). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 42*(2), 165-210. https://doi.org/10.1162/COLI_a_00237

Tao, C. C., Kim, S. K., Lin, Y. A., Yu, Y. Y., Bradski, G., Ng, A. Y., & Olukotun, K. (2006). Map-reduce for machine learning on multicore. *Advances in Neural Information Processing Systems, 6*, 281-288.

Topaloglu, U., & Yu, F. (2012). Towards large-scale Twitter mining for drug-related adverse events. In *Proceedings of the 2012 International Workshop on Smart Health and Wellbeing* (pp. 25-32). https://doi.org/10.1145/2389707.2389713

Tare, M., Gohokar, I., Sable, J., Paratwar, D., & Wajgi, R. (2014). Multi-Class Tweet Categorization Using Map Reduce Paradigm. International Journal of Computer Trends and Technology, 78-81

Tushar, R., & Srivastava, S. (2012). Analyzing stock market movements using Twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*. IEEE Computer Society.

Vimal, S., Robinson, Y. H., Kaliappan, M., Vijayalakshmi, K., & Seo, S. (2021). A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction. *The Journal of Supercomputing, 77*(1), 1-17. https://doi.org/10.1007/s11227-020-03268-0

Vimal, S., Robinson, Y. H., Kaliappan, M., Abdulellah, A. A., & Ashish, K. L. (2021). AI-based forecasting of influenza patterns from Twitter information using random forest algorithm. *Human-Centric Computing and Information Sciences, 33*, 1-14.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT and EMNLP* (pp. 347-354).

Wook, M., Yahaya, Y. H., & Wahab, N. (2009). Predicting NDUM students' academic performance using data mining techniques. In *Proceedings of the Second International Conference on Computer and Electrical Engineering* (Vol. 3, pp. 357-361). https://doi.org/10.1109/ICCEE.2009.24