Sentimental analysis on English Tweet on Demonetization and Analysis the effect of Demonetization of Digital wallet of India.

MSc Research Project
Data Analytics

Sakshi Khanvilkar
Student ID: x22117776

School of Computing
National College of Ireland

Supervisor: Abid Yaqoob

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sakshi Kishor Khanvilkar
Student ID:	X22117776
Programme:	M.Sc. Data Analyst **Year:** 2023-2024
Module:	M.Sc. Research Project
Supervisor:	Abid Yaqoob
Submission Due Date:	31/01/2024
Project Title:	Sentimental analysis on English Tweet on Demonetization and Analysis the effect of Demonetization of Digital wallet of India
Word Count:	10020 **Page Count** 24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sakshi Kishor Khanvilkar

Date: 31/01/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Sentimental analysis of English Tweets on Demonetization and Analyzing the Impact of Demonetization on Digital Wallets in India.

Sakshi Khanvilkar
x22117776

Abstract

On November 8th, 2016, the Prime Minister of India declared the demonetization plan India to address the issue related to "black money" and fraudulent notes. This unprecedented move declared 86% of the country's money illegal, explicitly all INR 500- and 1,000-rupee notes. While India is a cash-based economy, the unexpected move caused huge problems, including payment. Several studies have been conducted to analyze the impact of demonetization on digital payments. Despite having different views and opinions, part of the population was strongly offended by the decision. People used various social media platforms to express their worries and points of view, aiding in a borderline conversation about the effectiveness and consequences of this revolutionary act. The effect of demonetization needs to be analyzed to understand how it affected the digital payment system in India and how an individual reacts to such a sudden change in the economy. For sentiment analysis of demonetization tweets, models such as LSTM, CNN, and Naïve Bayes were proposed. Further, the model was compared based on accuracy, recall, and F1 score. The time series analysis was implemented on financial data collected from the database of the Indian economy. To observe the trend of digital payments in India's later demonetization, models like LSTM and exponential smoothing were implemented. For sentiment analysis, the LSTM model has given better results. The Time Series model namely: LSTM and Exponential Smoothing with lower MSE (mean square error), RMSE (root mean square error), and MAE (mean absolute error) is an optimal model for the prediction of the cashless economy trends in India.

Keywords: Demonetization, Time series forecasting, LSTM, CNN, Naïve Bayes, Exponential Smoothing.

1 Introduction

1.1 Background

Demonetization is defined as the way of removing a money unit's legal tender status and has been used by countries all over the world to address economic problems. We used a big demonetization strategy in November 2016, deleting large currencies of Rs 1000 and Rs 500. This program, led by India's Prime Minister, aimed to eliminate corruption, black money, and forgery. It created a debate, while this method was meant to enhance the money system and promote cashless transactions After a month of demonetization on altered domain of the Indian economy sector has become plain. This sector discusses the effects of corruption and illegal money as well as the issues affecting rural areas. This important move has received worldwide attention, with both favorable and negative responses, making it a topic of intense debate and scrutiny. The public is divided, with one side upholding the money prohibition, one side opposing it, and neither side being positive. India is a cash-based economy, with money accounting for most transactions. As per the data, total notes production as of 4th November 2016 was Rs17,742 billion, which represents 13% of GDP, whereas the value of Rs 500 and Rs 1000 constituted 86.5% of notes in production, allowing for 11% of notes in existence (Chopra, 2017). Since the replacement of currencies was lacking for certain periods, citizens had to start different ways of payment to carry out financial activities. With a large population having mobile phones, together with other variables such as internet access it became easier for banking and other financial services to provide digital payment modes which led to an increase in the cashless economy (Manocha, 2019). In the month post demonetization, the use of digital modes like NEFT, RTGS, Mobile transactions, and wallet companies like Paytm, PhonePe, and Google Pay display significant growth in terms of both value and volume. Nevertheless, after the enormous popularity of demonetization, the government of India started supporting the digital banking service, holding a target of 25 billion digital transactions for the financial year 2021 (Singh, 2017). On the other hand, social media networks have developed a platform where people can straight away examine their behavior and thoughts. Demonetization has become a major topic on the internet where people share their thoughts on Facebook and Twitter[1] (Vempaty, 2016). After the revelation made by the Prime Minister, the social network generated 650,000 tweets within 24 hours following the disclosure, and a million more in the week that followed. Studying tweets for a review of the impact of demonetization on many sectors while showing communication behavior. This research examined the long- and medium-term consequences of demonetization on India's banking system, as well as whether the demand for demonetization was merely an interim outcome of cash shortages. A time series model has been implemented to examine the course of digital payment through the period and to see if there was any gap during demonetization time. Using Time series models such as LSTM (Long Short Term-Memory) and exponential smoothing the study aims to Look into patterns, trends, and irregularities in the volume and value of online transactions and the models will be further evaluated using metrics like RMSE, MSE, and MAE. Furthermore, analysis is done on the sentimental data to display different patterns that are either negatively or positively related to demonetization will be done using LSTM and CNN models and based on the classification report will determine which model was able to classify the positive and negative reviews.

1.2 Motivation

The Government seems to be promising people that demonetization is a phase that will transform the economy into a cashless economy. As per the World Bank Global Findex Data, only 4% of the total population (aged 15 and above) in India stated that they were using a bank account to collect salaries in 2014. Rural areas share even lower at 2.96%, with only 1.7% of persons in the lowest 40% households related to utilizing a financial account to get paid the prior year. These demonstrate that cashless transactions are possible. These display that cashless payments make up a small percentage of total transactions in the Indian economy (Nithin, 2019). In 2019, India faced an economic recession attributed, in part to demonetization among other factors. This complicated network impact raises the question of whether India’s demonetization has an ongoing effect on the spread of digital payment (Ohlan, 2019). The need to investigate this topic and use of time series model for analysis comes from the desire to figure out whether demonetization worked as a drive for a permanent shift in the digital payment model. The study also aims to collect insight into public reactions to such policies by considering people’s reactions to this economic policy. This method is not only for the demonetization event but also as an initial basis for expecting

[1] Twitter was recently re-branded to “X” by Elon Musk.

possibly the same reaction in a similar scenario, resulting in a larger understanding of public opinion and behavior toward economic change.

1.3 Research Questions

Has India's execution of demonetization had a substantial and long-lasting effect on the rise of digital payment, and how could sentimental analysis assist in understanding the public view about demonetization?

The two sub-research issues directed in the study are as follows:

SUBRQ1: How well does deep learning-based sentimental analysis determine public sentiment towards demonetization, collecting sentiments of support or opposition and determining whether public opinion is positive or negative?

SUBRQ1: Implementing a time series model of whether the spike in electronic transactions an intermittent result of the lack of currency during demonetization is or is a long-term effect.

1.4 Research Objective and Contributions

To address SUBRQ 1 and SUB BRQ 2, this research work has the following potential contributions:

- Investigate the consequences of demonetization on digital transactions.
- Evaluate if the growth in digital payments was the short-term effect of the money trouble during demonetization or is a long-term effect.
- Analyze the sentiments and opinions conveyed by the public concerning demonetization.

The remaining section of this article is planned as follows:
Section 2 depicts an overview of the literature as well as a significant analysis of the present literature examinations in the subject zone. It also elaborates on the diverse paths implemented to undertake data breakdown in the context of an outer event. This forms the basis for the research carried out on this project. Section 3 describes the study practice and the stages to be taken. Section 4 is the design section of the project where a detailed explanation is given of the project architecture. Section 5 establishes the operation of the project, followed by an evaluation of all the models implemented and discussed in detail. Section 6 depicts the results comparison and review of the complete study. Section 7 is the final section that closes the study indicates the restrictions of the project and recommends forthcoming investigation paths.

2 Related Work

The Division impact of demonetization represents an important analysis of the techniques deployed in prior studies to observe the impact of demonetization on India's payment system. Further, an examination of sentimental analysis is done using deep learning and machine learning models, with a focus on the public view on demonetization. The part also highlights the use of LSTM, CNN, and Naïve Bayes in prior studies presenting a short understanding of their function in sentimental analysis during the demonetization period. Additionally, the use of the time series analysis on time-dependent data when external incident or involvement happens. It highlights how the approaches mentioned in the paper are applicable in different scenarios to access effectively.

2.1 Impact of Demonetization

(Chavali, 2019) summarize a comprehensive analysis of Reserve Bank of India statistics from March 2015 to July 2018 focusing on the 20 months before demonetization. The study found an increasing trend in digital transactions following demonetization especially through Real Time Gross Settlement (RTGS) and Retail Electronic Clearing, mobile banking, and debit cards. The results strongly suggest that demonetization was an important factor in the uplifting of digital payment in India. The study uses various statistical tools, such as descriptive statistics for a short description of vital data aspects, regression analysis to find correlations, t-test to determine mean differences, and trend analysis to discover patterns over time. Also, the decision cost the Reserve Bank of India into profit of Rs 130 million. The study says that although a reduction in fake currency and an increase in digitization there was a negative impact, like the loss of 1.5 million employment, negative impact on real estate and other sectors. This study extends our insight into the impact of demonetization on the digitization of India by focusing on distinct columns implying digital transactions. This analysis adds to the literature by providing varied opinions on how columns like Real Time Gross Settlement (RTGS), Retail Electronic Clearing, mobile banking, and debit cards act to the change in economic policy. Since the study consists of secondary data from RBI limits the extent to which user preferences and patterns may be understood. Additionally, the study focuses on 20-month period data of post-demonetization which might not catch a full variety of long-term effects.

(Bansal, 2018) concise description of the use of electronic banking services following the effects of demonetization, with emphasis on the transformation of electronic transactions. The study used the t-test method to analyze the enhancement of digital banking services in both volume and value estimates during the pre-demonetization and post-demonetization periods. The dataset considered for this analysis was taken from the RBI (Reserve Bank of India) website and the period of the dataset was from November 2015 to October 2017. The t-test outcome for both volume and value indicates that demonetization had a substantial influence on the usage of digital payment modes. The volume-wise analysis states that digital banking services like RGTS and mobile Banking discovered a known value of less than 0.05. This indicates that there was a statistical increase in digital payment modes. Despite the t-test findings showing a distinction in the statistical characteristics of the two samples, the study fails to confirm that it fulfills the t-test's fundamental assumptions that the data must be distinct from one another. The dataset is made up of entries throughout time, hence the t-test results may be erroneous.

(Khatik, 2018) explores the growth of payment in India especially in the post-demonetization time. The research states that there was a significant growth in transactions via networks such as NEFT, CTS, IMPS, and NACH. The observed evaluation in this paper states that the country's swift change to a cashless economy, through digital way increased significantly between December 2016 and January 2017. Still, after this growth the next month was easy. A statistical examination of the next few months' payment data could have provided a better consideration of the impact of demonetization.

(Chopra, 2017) describes the short-term cost and medium-term gains of demonetization in various sectors. The study cites the increase in digital payments which lead to the move towards the cashless economy. The implementation of the Bharat Interface for Pay Money (BHIM) is particularly important. From November 2016 and February 2017, there was a certain amount of growth in both the number and value of digital payment transactions. The outcome of the study shows that there was a high increase in UPI Transactions, which reached 4.2 million in Jan-Feb'17, up from 0.3 million in Nov'16. Although the study discloses that demonetization increase digital payment in India, it doesn't show what happened in subsequent decades.

Focusing on NEFT, RTGS, and mobile banking during the pre- and post-demonetization, (Hindocha, 2019) illustrates the analysis of electronic payments. The research was performed using the RBI annual report which consists of data from different govt organizations. The paper states that there was a significant growth in the use of digital payments in Feb'17 compared to Nov'16, looking to various payment modes like National electronic fund transfer (NEFT), cheque Truncation system (CTS) Immediate Service Payment Method (IMPS) and National Automated Clearing House (NACH) the paper states that IMPS give high as 196.7% in January'2017. Other payment methods such as NACH spiked by 116.7%. The report does demonstrate that electronic transactions rose following demonetization, but it does not detail what happened in the decades that followed. summarizes the impacts of demonetization on payment modalities such as cellphone transactions, POS transactions, card transactions, and so on.

(Nithin, 2019) The study implies time series analysis and outcome showcase that the use of mobile transactions had increased compared to card transactions and point of sale transactions. The paper uses two time series analyses, the ARIMA model and the Intervention time series analysis. The results of this model indicate that demonetization had an uncertain impact on digitalization.

The above studies demonstrate an immense rise in electronic payment options like RTGS, Mobile Banking, UPI, NEFT, and other papers suggesting exciting changes towards a cashless economy. The research demonstrates demonetization acts as a catalyst for the rise in digital transactions.

2.2 Sentimental Analysis of Demonetization Tweets

The following paper discusses different approaches and responses in statistical and sentimental analysis.

(Kumar, 2019) surveyed public opinion during the demonetization period in 2016. The research was carried out by implementing different steps including data gathering, pre-processing, and sentiment analysis. Following a review of 3,000 sample tweets, some tweets unfavorable towards demonetization had an average sentiment score nearing zero stating that sentiments were negative but not excessive. The geographical distribution of tweets discloses most contributors in Bangalore, India. Around 92% of tweets were written in English followed by Hindi tweets of 2.73 % and regional language 5.27%. The paper refers to the use of Text Blob, a Python library with a Natural language toolkit (NLKT), for processing and analysis of tweets. The author highlights challenges with text Blob such as inconsistencies in sentimental computation for certain tweet words. Although the study used the text blob as a new approach, it failed to specify the usage of a sentimental analysis model other than text blob.

(Singh, n.d.) shows an examination of feelings in code-combined texts and categorizes phrases as angry, sad, or cheerful. The Twitter website and its live streaming service were utilized to collect and assess the dataset for this investigation. For this experiment, Nave Bayes with character n-grams, Nave Bayes with word n-grams, LSTM, and support vector machine were used. The model performance was further assessed using statistical measures such as F1 score, recall, precision, and accuracy, and it was discovered that subworld LSTM provided 71% accuracy. As a result, the accuracy of this research may be increased by improving the dataset and attention-based model. Tweets written in English and Punjabi related to agriculture are conveyed in this

(Goyal, 2019). The paper uses data from Facebook, Twitter, and YouTube. To make statements readable and meaningful the author removed all the emoji. The author further shows the comparison of Naïve Bayes with Support Vector Machine (SVM) and differentiates between positive, negative, and neutral categories.

(Dar, 2017) experimented with sentimental analysis on tweets and retweets associated with the government policy of demonetization, the author considers a dataset having 8000 tweets and 6 features from the Twitter website. The research is carried out by deploying a combination of lexicon-based techniques and machine learning models such as Support Vector Machine (SVM), Classification and Regression tree, Random Forest, Logistic Regression, and KNN. Among all classification models, SVM (89%), CART (90%), and logistic Regression (88%) were established to be the most effective combination.
The following (Goularas, 2019) provide useful assistance for deep learning methods. The analysis of the sentimental is carried out through different modes like conventional Neural Networks (CNN), Long short-term memory (LSTM), and word embedding. Out of all the methods, word embedding was found to be the best.
(KOLYA, 2020) performed a sentimental analysis of public feedback on India's Goods and Service Tax delivery in 2017. The author gathered nearly 2,00,000 tweets related to the GST and implemented a relevant sentimental model to analyze the tweets. Sentiment words were identified and granted polarity scores using the Naïve Bayes bag of word algorithm. The study used an LSTM model to predict the sentiment and overall accuracy of 84.51%. However, the analysis relied just on English words, possibly missing out on vital information from tweets in other languages.
(Srinivas, 2020) impacts the domain of sentimental analysis by researching the use of deep learning models like LSTM and Neural Networks. The author of both models concluded that the LSTM model is better as it gives an accuracy of 87%. The dataset used for the research was taken from the Kaggle consisting of 1.6 million tweets.
(Harish, 2018) illustrate the sentimental analysis done on the Twitter data taken from November 9 to December 3rd and performed machine learning models like Naïve Bayes, SVM, and Decision tree out of all the models SVM performed well with an accuracy of 84%. The study on sentimental analysis of Twitter data related to demonetization indicates an ongoing positive sentiment.
Ultimately the selected research paper above collectively gives important insights into sentimental analysis during events like demonetization or implementation of the Goods and Services tax in India. The following papers use an array of algorithms to estimate public opinion from social media data, including TextBlob, Naïve Bayes, Support Vector Machine, LSTM, and neural networks. The above studies also highlighted the challenges while working on code-mixed text, and language differences.

2.3 Time Series Analysis on Digital Payment Data.

It was only three to four years ago that cashless payment options like digital wallets and transfers became widespread. The goal is to look at how the learning approach might be utilized to better comprehend the effects of demonetization on India's cashless economy. The articles that follow discuss various deep-learning approaches and strategies.
(Siami-Namini, 2018) have conducted a comparison of statistical approaches such as Autoregression Integrated Moving Average (ARIMA) with deep learning techniques such as long-term short memory (LSTM). The study results demonstrate that the deep learning approach LSTM outperformed the statistics model. In contrast with the ARIMA model, LSTM has an average error rate reduction of 84% to 87% showing better forecasting accuracy. The author also specifies the number of training iterations designated as epochs. The study also shows that an easier technique with epoch = 1 may nevertheless predict accurately in the case of rolling forecasting, even though increasing the number of epochs does not necessarily give better results. In terms of assessment criteria, the study falls short of providing a summary of statistical measures such as F1 scores, recall, or the ROC curve, which might highlight the model's effectiveness while dealing with an unbalanced dataset.

The (Pramod, 2023) represents the comprehensive review of currency forecasting using the RBI dataset of the SENSEX Index, and NIFTY 50 index, the analysis is done using different forecasting models like deep learning and machine learning model. The Ensemble ML model excelled the Deep learning model earning 90% accuracy in prediction. The paper highlights the inconsistent accuracy of the Deep learning model, noting the difficulties of using complex neural networks for financial forecasting. In the end, the study adds greatly to the field of exchange rate prediction using machine learning models. However, for more deeper analysis the author should analyze the clarity of complex models and development needed in the deep learning model.

(Asati, n.d.) demonstrate a study estimating the Indian consumer price index. The writer has used the LSTM model to create several time series models, including XG boost, statistical learning (ARIMA, Theta, prophet), and deep learning. The Ministry of Statistics in India provided the dataset that was used for this investigation, which covered the period from January 2011 to December 2021. Mean absolute deviation (MAD), Root Mean Square Error (RMSE), and Absolute Percentage Error (MAPE) were used to assess the model's performance. The Prophet model shows less MAPE (0.44), showing its flexibility in handling the outliers and missing data whereas LSTM shows the MAPE (0.18) showing its capabilities to handle complex patterns in the time series dataset. ARIMA and Theta model gave equal MAPE (1.48) signifying similar outcomes but a larger error than prophet and LSTM. Still, the study has limitations such as including its dependence on univariate analysis and potential data gaps. To enhance the prediction accuracy further development could include a multivariate approach.

(Pirani, 2022) examines the performance of stock price performance of different Banks and estimates the performances of the LSTM model, Bi-Directional LSTM, and Gated Recurrent LSTM using RMSE. The overall results show that GRU consistently accomplished well in terms of accuracy and speed compared to Bi-LSTM and LSTM. The study also discovered that Bi-LSTM improved forecast accuracy by 11.86% on average and that using GRU with one gate less than LSTM-based models improved accuracy by 20.32%. Although the researchers have touched on the behavioral analysis of the model still deeper, investigating the specific findings produced from the model and their use in decision-making could have significant value.

The following (Dr Jai Kishan Karahyla, 2023) offers a hybrid model for predicting Bitcoin's price direction, the study is carried out using LSTM and Dense Neural Network (DNN). Before processing with implementation of the model author has done some preprocessing steps like normalization, and filling missing values, feature extraction using Principal Component Analysis (PCA). Model performances are evaluated using statistical methods namely accuracy, precision, and recall. The LSTM model proved an accuracy of 89.43% with precision and recall of 86.5% and 83.44%, whereas the DNN model gave an accuracy of 95.34% followed by precision and recall of 92.49% and 90.28%. The author mentioned that the hybrid LSTM-DNN model has an accuracy of 97.7% along with a precision of 95.23% and recall of 92.56% which is higher than the accuracy of the individual two models. (Chamsukhee, n.d.) represents the forecasting model using Exponential Smoothing, Long Short–Term Memory (LSTM), and ARIMA model. The MSE received for models Exponential Smoothing, was 0.150 for LSTM 0.127 and for ARIMA 0.153. Furthermore, the accuracy of forecasting may vary depending on activity or change in the E-Database. Despite all the limitations, the outcome indicates that LSTM is a better prediction model.

Overall, the studies for evaluation show the dominant role of Long Short-term memory (LSTM) which has enhanced the precision of the time series problem across different financial domains. The papers also illustrate the LSTM advantage of handling complex patterns and improving prediction precision over the traditional statistical mode

3 Research Methodology

For this study, the aim is on two key tasks: sentimental analysis of tweets associated with demonetization and time series forecasting of digital payments in India. The objective of the sentimental analysis is to unfold public views and opinions stated on social media during the demonetization phase. Similarly, analyzing trends and patterns in digital payments in India is a component of time series forecasting where I am analyzing the changing patterns of digital payments transactions by using historical data. The study aims to afford a thorough view of the impact of demonetization on public attitude and following trends in digital payment use by incorporating sentimental and time series analysis.

3.1 Overview of KDD

In the background of the problem statement, it is essential to acknowledge the different methods for leading data analysis projects present, each with its own method. Among the available methods are KDD (Knowledge Discovery in Database), CRISP-DM (Cross-industry standard process for data mining), and SEMMA (Sample, Explore, Modify, Model, and Assess). According to the research topic, this project comes under the KDD methodology. Figure 1 explains the idea of knowledge invented in the database as a way of searching, cleaning, and processing substantial data from raw databases to identify the hidden patterns and analyze the outcome to capture necessary insights.

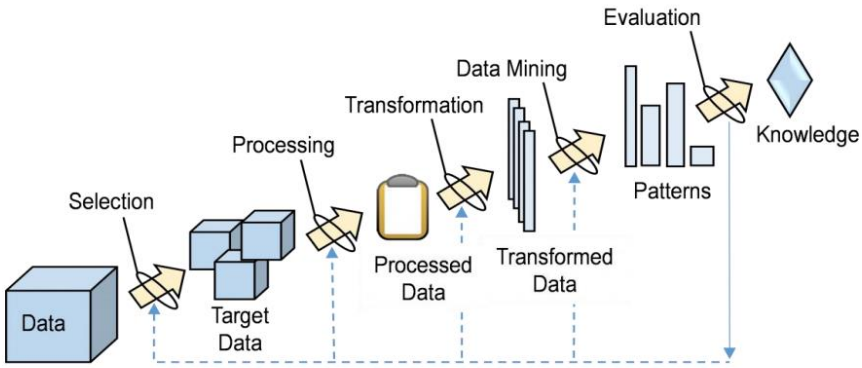

Figure 1: KDD Architecture (ref: Alam, 2023**)**

3.2 Data Gathering

The primary goal in the Data collection part of the KDD method is getting full insight into present data and collecting the necessary information to support future analysis. For this study, two different datasets are considered. The Twitter dataset is taken from Kaggle[2] and it consists of 14,940 records having a time cover between November 2016 and April 2017. For this research analysis, taking a sample of tweets because of a limited period, the dataset is gathered with keywords such as '#Black money', 'Demonetization', and '#Modi'. The study focused on evaluating general acceptances of government policy and analyzing public opinion as an outcome of the demonetization program**.** Another dataset for the time series analysis is taken from The Reserve Bank of India's (RBI) database of the Indian economy[3]. From April 2004 to October 2019, the dataset comprises the monthly value and volume of financial transactions conducted via various digital payment mechanisms. This data is chosen because it spans a large valuable interval before demonetization, allowing for analysis of not only the immediate but also the medium- and long-term implications of the period. Using the

[2] https://www.kaggle.com/datasets/arathee2/demonetization-in-india-twitter-data
[3] https://rbi.org.in/Scripts/Statistics.aspx

LSTM model, the author built numerous time series models such as XG boost, statistical learning (ARIMA, Theta, Prophet), and deep learning. The data used in this analysis came from India's Ministry of Statistics and ranged from January 2011 to December 2021. The model's performance was assessed using Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Deviation (MAD).

3.3 Data Pre – Processing

This step of KDD focuses on increasing the quality and validity of the data providing it is well-prepared for the analysis to come. Two datasets are selected for the project development; both the data are collected in raw format. Therefore, to make it more accessible and understandable both datasets have undergone the preprocessing stage.

The first dataset which consists of demonetization tweets is downloaded from Kaggle. The downloaded demonetization tweet dataset had many columns like 'text' which includes the tweet text, favorited, favoriteCount, created dates, retweet, and truncated. Still, for this study, only the field 'text' which contained all the tweets is considered. The tweet dataset entails data like URLs, misspellings, unwanted punctuation, Hashtags, Emails, Person names, Company Names, and sentiments hence want series of preprocessing steps to be included like Special characters removal, Stop words removal, Tokenization of the text column in the dataset **(**Unnithan, 2022). The second dataset which is the monthly financial transaction is downloaded from the RBI database of the Indian economy, the preprocessing of the financial dataset includes, reviewing and understanding missing values of the dataset, and performing a thorough examination to identify and rectify any data gaps. Also filling the null values with zeros to have a consistent and broad dataset.

3.4 Data Transformation

Many modifications have been made to the dataset variables to make the process easier to comprehend and effective. In the domain of sentimental analysis, a dataset consisting of 14,490 records of November 2016 and April 2017 has been thoroughly cleaned. The text column that defines the actual tweets has been modified using regular expressions from Python Libraries to maintain consistency and can be used in later analysis. The key aspect of this dataset is the addition of a sentiment column, in which the labels are added according to the sentiments using the Text Blob library of Python. Similarly, the financial dataset which consists of the 'Month Year' column representing the dates in a format like 'April 2004' is formatted to a consistent presentation like '2004-04-01'. Also, the features in this dataset are scaled to make better comprehension and consistency.

3.5 Data Minning

Data mining is considered the major procedure in most cases of data analysis. In this the stage of data mining process, the provided data is taken as input and endeavors to generate the essential output[4]. In this study, for sentimental analysis of demonetization tweets, three models are used namely: Long Short-Term Memory (LSTM), Conventional Neural Network (CNN), and Naïve Bayes. The sentimental analysis using LSTM has shown better results like (Goyal, 2019). Thus, by taking it as the base model, further comparison has been shown by implementing CNN and Naïve Bayes model for the Classification of Positive, Negative, and Neutral responses of demonetization tweets. The dataset is split into training and testing with 80% of the data used for training and 20% for testing. For the Time series analysis of financial data, Long Short-Term Memory (LSTM), and Simple Exponential smoothing

[4] https://www.analyticsvidhya.com/blog/2021/05/introduction-to-data-science-project-lifecycle/

(ESE), two models are applied. The dataset is divided into training and testing, where 80% of the data is used for training and the remaining 20% is used for testing. In contrast to prior studies, the forecasting of financial data is done using T-test (Bansal, 2018) and ARIMA. By employing the LSTM network, my study takes a distinct approach. The outcome obtained from the LSTM model for time series problems demonstrates significant improvements in forecasting accuracy and insight building when compared to the previous research methods.

3.6 Evaluation

The last stage of KDD architecture is Evaluation, the performance measures for the sentimental analysis model involve examining the accuracy, precision, recall, and F1 score. This metric reveals the model capacity of how it correctly classifies the emotions of the public regarding demonetization. In time series analysis, the Mean squared error (MSE), Root mean Squared error (RMSE), and Mean Absolute Error (MAE) are used to analyze the accuracy of the model.

4 Design Specification

The objective of this research is to establish an accurate model for analyzing the opinions of the public in a set of tweets about demonetization and to understand trends in India's digital payment system after the demonetization phase. Data gathering, data cleaning, data pre-processing, exploratory data analysis, feature extraction, model training, prediction, and Evaluation are all the critical stages in the presented framework. This research has two approaches: firstly, it category the emotions to analyze public reviews on demonetization. Secondly, predict changes in financial data using time series analysis. Every stage in this extensive structure is highlighted in depth admitting a comprehensive understanding of the research aim.

4.1 Sentimental Analysis of Demonetization Tweets

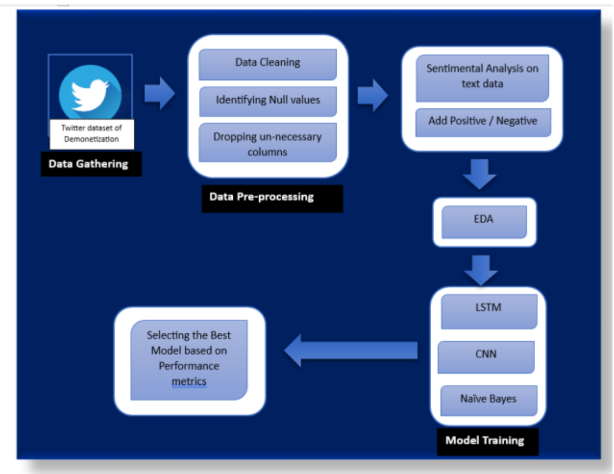

Figure 2. Design Architecture for Sentimental Analysis

4.1.1 Dataset Understanding

For this study, data is collected from Kaggle for analysis of the decision to demonetization which was made by the Indian government. To eliminate the black money was the major goal of implementing the demonetization. In November 2016 the demonetization program was introduced, the initial days of demonetization were like emergency days in India. However, the government the important steps required for demonetization. All the sentiments of Indian people were considered for examination (Darliansyah, 2019). The dataset for demonetization tweets consists of 14,490 records, holding sentiments and responses during the epoch from November 2016 and April 2017. It includes features like 'text', 'favorited', 'favorite Count', 'replyToSN', 'created date', and so on. The text column, which contains the content of tweets, and the created date column which states at which date the tweet was created is the key focus of the features.

4.1.2 Data Pre-processing

Data treatment is a vital step to remove noise from the dataset to perform better analysis. The sequences of data processing are utilized for data cleaning and null value removal. Some of the features from the Twitter dataset are not useful for analysis such as 'favorited', 'favouriteCount', 'replyToSN', 'truncated',' retweetCount',' retweet,' retweeted', and so on. Therefore, the initial step is to filter out the columns and then replace the name of the column consequently. Moreover, I rectify, if there are any null values in it, but it shows that there were no null values in it. After all these steps, the date column format is changed into a date-time format to fetch only dates from the columns. Next step, I clean the text column which consists of URL, Hashtags, unwanted punctuation, @, replacing words for example #demonetization is replaced with demonetization, @Modi is restored with Modi. To preprocess the text column certain steps were implemented like Special characters removal, Stop words removal, and Tokenization of the text column in the dataset. This complete preprocessing step is done using Python.

Special characters Removal: Regular expression (Regex) libraries are used to filter out the special characters like URL, hashtags #, @, and username for example #Demonetization is restored with demonetization or @Modi is restored with Mode.

Stop word Removal: For this python, NTLK's stop words are used to clean the data for this study. Following the preprocessing a new column was added named 'cleaned tweets.

Tokenization: This is a method of splitting down a sentence into individual words or terms called a token. Keywords have been taken from the cleaned tweets column of the dataset. The split function of the tokenization library is used to separate words in the connected text and the counter function is used to get the frequency of the words. Therefore, a list of the most accepted words used in tweets on demonetization was developed.

4.1.3 Feature Engineering

In the sentimental analysis of Twitter data, The Text Blob library of Python was considered to assess the polarity of each cleaned tweet column in the dataset. A function was created in Python, that allocated a sentimental tag to each tweet based on polarity score. The sentiment label was 'mod_pos' as moderate positive, 'high_pos' as extremely positive sentiments, 'mod_neg' for moderate negative sentiments, and 'neutral' for balanced sentiments.

4.1.4 Exploratory Data Analysis

The Exploratory data analysis stage comprises an in-depth review of sentimental analysis in the demonetization Twitter dataset. The sentimental labels are visualized through bar graphs and pie charts, providing useful insight into the overall sentiment content of the demonetization Twitter dataset refer to Figure 3.

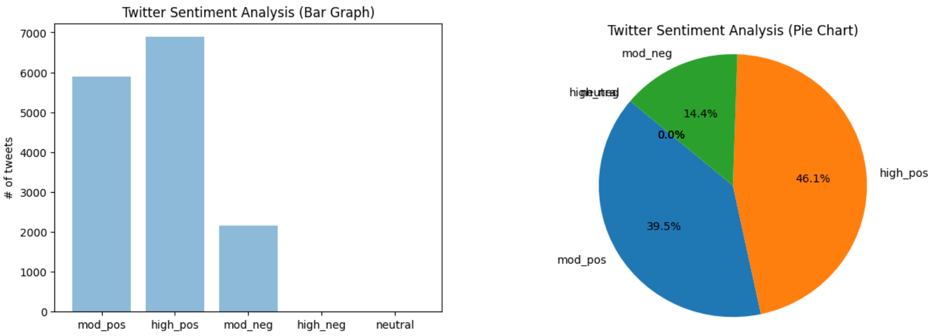

Figure 3. Sentimental Labels Visualization

Examine the bar graph, most of the tweets come under the moderately positive and strong positive sentimental classes, representing an overall positive outlook in the discussion. Particularly, the number is moderately negative is very low. The pie chart illustrates the percentage distribution, with 46% of tweets expressed extremely favorable to demonetization, followed by 39.5% implying moderately good expression.

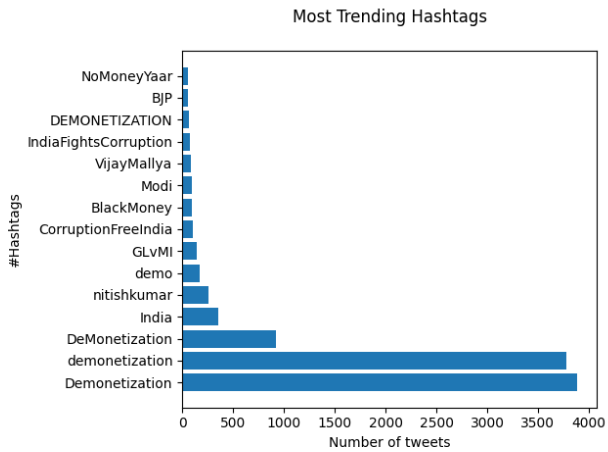

Figure 4. Frequency of Trending Hashtag words

Figure 4 shows the measures of the number of times each hashtag has appeared and identifies the top 15 most used hashtags in the demonetization tweet dataset. The horizontal bar chart shows the popular hashtags whereas the x-axis shows the frequency of each hashtag's events.

4.1.5 Model Implementation

In this work, I employed two strategies to categorize positive and negative sentiments: Deep Learning and Machine Learning. Each approach uses 80% of the data sample as training data and the remaining 20% as a test dataset. LSTM and CNN are the deep learning algorithms employed, whereas Nave Bayes is the machine learning algorithm.

LSTM Model

The LSTM model is an artificial memory architecture to stimulates the function of the brain's memory system and language understanding. As per psychologists and neuroscientists, emotions play a vital role in multiple phases of the memory process such as keeping information, cementing memories, and recovering experiences. For example, an emotional state at the time a person experiences and analyses an observation can affect the encoding of data into short-term or even long-term memory (Unnithan, 2022). The LSTM model built for this study has different layers and every layer has its contribution. The first Embedding layer of the model is mostly used for processing sequences. It is responsible for converting positive integers onto dense vectors. The SpatialDropout1Dlayer forms the second layer responsible for setting randomly an amount of input units as 0 at each update all over training time. This layer also prevents overfitting. The third layer forms the LSTM layer consisting of 176 units. Last the Dense layer with 2 units and a SoftMax activation. In sentimental analysis, using SoftMax activation helps to categorize a piece of text (such as text, or review) into one or several sentimental classes (for example positive, negative, neutral).

CNN Model

The CNN model can obtain a set of variables from the global dataset and can create a link between those variables. This makes the CNN model more suitable for classification problems and gives better accuracy. In natural language processing, the text of the sentence can be collected piece by piece and piece and the link between each of these is considered, without considering the whole text the sentiment might be understood wrong. CNN model has a convolutional layer to collect data from the larger context, therefore CNN can be used for sentiment analysis by creating a simple convolutional neural network model (Huang, 2022). The CNN model implemented for this research has several layers like the Embedding layer that convert the word scores into compact vectors, which is necessary to represent the meaning of words. The following Permute layer transposes the dimension of input which is further given for convolutional processing. The model also consists of two convolutional layers holding a 'real' activation function with filter units responsible for identifying a local pattern in the input text. The maxPooling layer followed by the GlobalMaxPooling1D layer extracts maximum value along the time dimension, reducing information into a fixed size vector. The final Dense layer produces output for classification.

Naïve Bayes Model

It is a set method for classification algorithms depending on Bayes Theorem. When applied to text data analysis, The Naïve Bayes model gives better results. To perform the sentimental classification, it considers the concept of mixture models. The Mixture Model is skilled in implementing the possibility of the component that it consists of the Bayes theorem to perform as a classifier based on probabilistic (Singh, n.d.).

4.1.6 Model Evaluation

Different metrics were used to analyze the performance of every model while finding the best model for the sentimental analysis of demonetization tweets. Notable metrics such as accuracy, precision, F1 score, and recall mentioned in Table 1 were applied since the sentimental analysis is named as a classification problem. The following statistical metrics were used.

Accuracy: It calculates the ratio of correctly predicted appearances to the total occurrences to represent the model's overall accuracy. In the context of emotional analysis accuracy identifies how effectively the model predicts sentiments throughout the classification.

Precision: It is defined as the proportion of true positive predictions to the total number of true or false positives. Precision in the context of sentimental evaluation examines the accuracy of the model's favorable predictions, demonstrating how trustworthy the model is when forecasting a positive sentiment.

Recall: The ratio of true positive predictions to the sum of true positive and false negatives is measured by recall. It is beneficial in emotional analysis as it evaluates the model's ability to identify the presence of positive sentiment, referring to its sensitivity to positive cases.

F1 Score: The Harmonic average of precision and recall is exploited to get the F1 scores. It provides an even evaluation of the model's performance when the class is not distributed evenly. The F1 score is useful for identifying trade-offs relating to precision value and recall value.

TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

Parameter	Formula
Accuracy	(TP+TN)/(TP+TN+FP+FN)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
F1 Score	(2*precision*recall)/(precision+recall)

Table 1: Metric Formula

4.2 Predicting Digital payment in India based on Time-Series Analysis

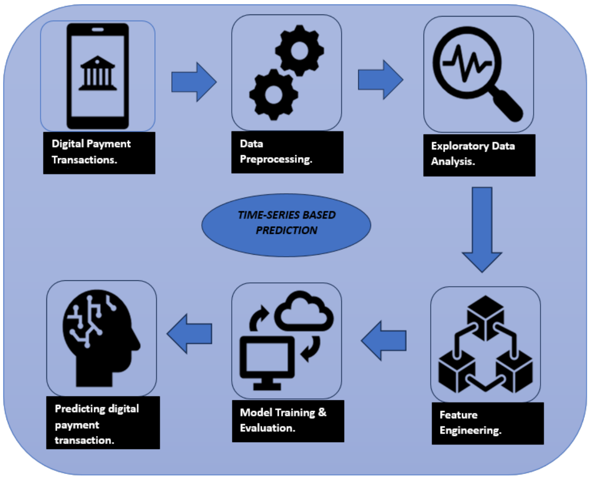

Figure 5. Design Architecture for Time Series Analysis

4.2.1 Data Understanding

The Reserve Bank of India (RBI) database of the Indian economy was used to compile the financial dataset. The information includes the monthly value and volume of digital payment transactions from April 2004 to October 2019. The dataset was chosen because it spans a large period before demonetization. This enables us to recognize not just short-term but also medium- and long-term occurrences of that duration. The data set consists of 187 records that show the monthly value and number of transactions. This study considers Real Time Gross Settlement (RTGS), Retail Electronic Clearing (REC), debit and credit card usage at POS, and mobile banking, all in terms of value and volume.

4.2.2 Data Pre-processing

In the data processing stage, the first step comprised converting the "Month Year" column into the standardized date format using the Python datetime library. After that, the dataset was reviewed for null values, and I replaced those with zeros. The dataset was further limited to a distinct event window of 35 months that is pre-demonetization (January 2015 to November 2016) and 35 months post-demonetization period (December 2016 to October 2019). The objective of narrowing the period was to analyze and inspect any possible impact of demonetization on digitization patterns in the period shown. The Financial dataset was scaled to get a better understandable illustration of values. The values columns comprising Real Time Gross Settlement (RTGS), Retail Electronic Clearing (REC), debit and credit card usage at POS, and mobile banking have been reduced from millions to billions. Similarly, the value columns consisting of Real Time Gross Settlement (RTGS), Retail Electronic Clearing (REC), and debit and credit card usage at POS, and mobile banking were tuned into values conveyed in millions. Furthermore, two datasets were created, one representing digital transactions in volume and another representing digital transactions in value.

4.2.3 Exploratory Data Analysis

While analyzing the data, the following stage is to do exploratory data analysis to get findings such as patterns, seasonality, and any visible output rapid change in the series. Figure 6 (a) depicts the volume of transactions, whereas Figure 6 (b) depicts the value of individual transactions from March 2014 to October 2019, with the date of demonetization represented by a red dashed line. The number of digital transactions grew significantly after the demonetization era, although the gap in transaction value is not visible for all payment mechanisms.

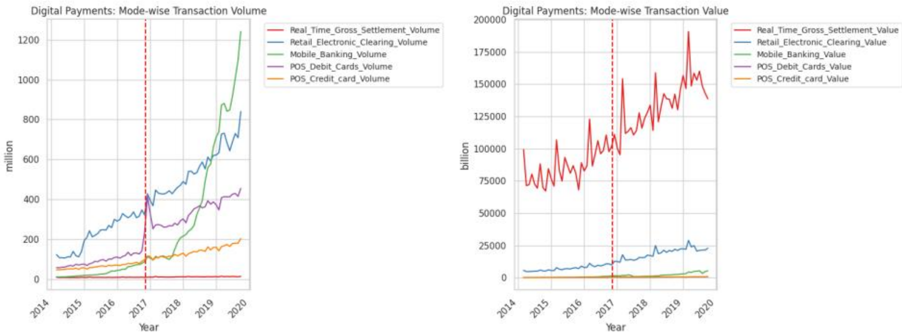

Figure: 6(a) Digital payment in terms of Volume; (b) Digital payment in terms of Value

The following Figure 7 graph depicts the transaction value and volume statistics, evaluating the variables such as mean, maximum, and minimum between pre- and post-demonetization. This representation clarifies the dataset's time behavior giving insight into the impact of demonetization on transaction patterns.

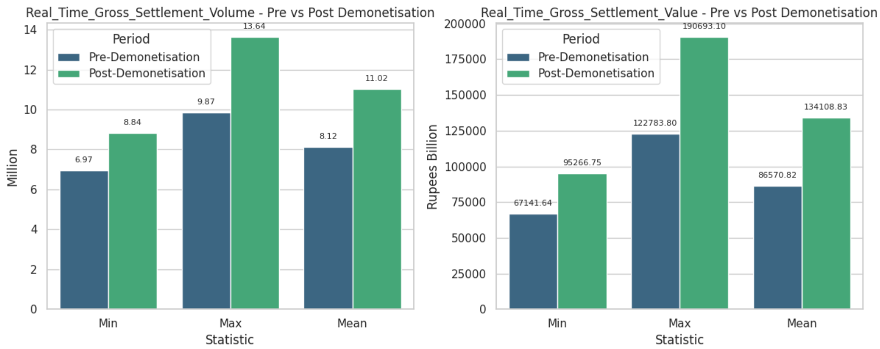

Figure 7: Real Time Gross Settlement (RTGS) in pre- and post-demonetization

Figure 7 illustrates the statistical information of RTGS transactions that occurred in the 34-month window pre-demonetization and 34-month post-demonetization in the relationship of value and volume. In the post-demonetization period, the minimal RTGS transaction volume varied between 6.97 to 8.84 million. Whereas the post demonetization period the overall RTGS Value ranged from 9.87 to 13.64 million. The lowest RTGS transaction value ranged from INR 67141.64 to INR 95266.75 billion and the highest RTGS transaction value ranged from INR 122783 to INR 190693. The average predicted volume of the mean is 13.63 million and the average predicted value of the mean is INR 2,20,200.

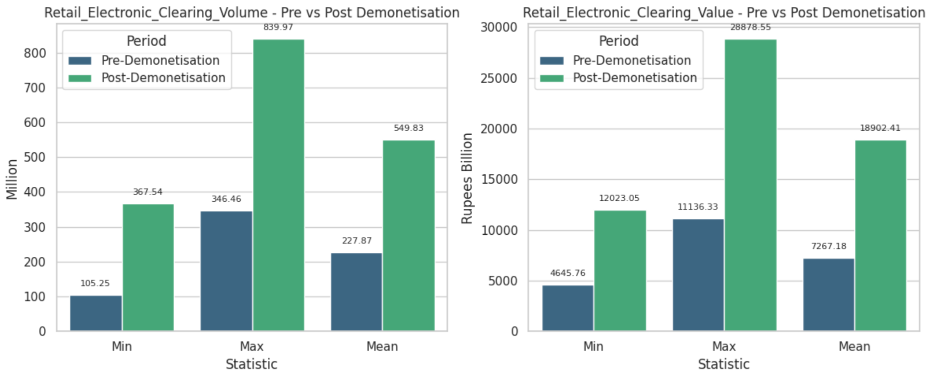

Figure 8: Retail Electronic Clearing Volume (REC) in pre- and post-demonetization

Figure 8 shows the usage of the REC model in terms of volume and value that happened in a 34-month window pre-demonetization and 34 months post-demonetization. In post demonetization, the minimum REC volume ranged from 105.25 to 367.54 and the highest REC volume varied from 346.46 to 839.97. The minimal value transaction of REC in post demonetization ranged from INR 4645.76 to 12023.05 and the highest REC transaction value ranged from INR 11136.33 to INR 28878.9. The average predicted volume of the mean is 514.86 million and the average predicted value of the mean is INR 16,638.

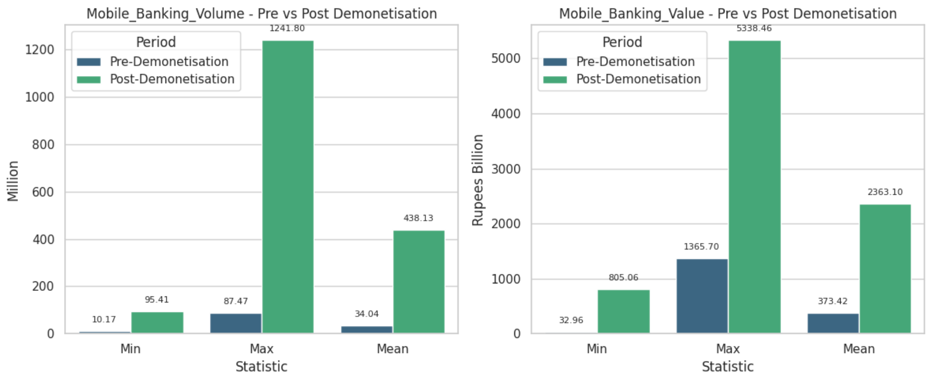

Figure 9: Mobile Banking in Pre and post Demonetization

Figure 9 depicts descriptive statistical information on Mobile banking that happened in a 34-month window pre-demonetization and 34-month post-demonetization in the relationship of value and volume. In the post-demonetization period, the lowest Mobile Banking transaction volume changed between 10.17 to 95.41 million. Whereas in the post-demonetization period, the overall Mobile Banking Value ranged from 78.12 to 1241.8 million. The minimum Mobile Banking transaction value fluctuated from INR 32.96 to INR 805.06 billion and the utmost Mobile Banking transaction value is from INR 1139 to INR 5338. The mean predicted volume of the mean is 251.4 million and the average mean value of the mean is INR 1,523.9

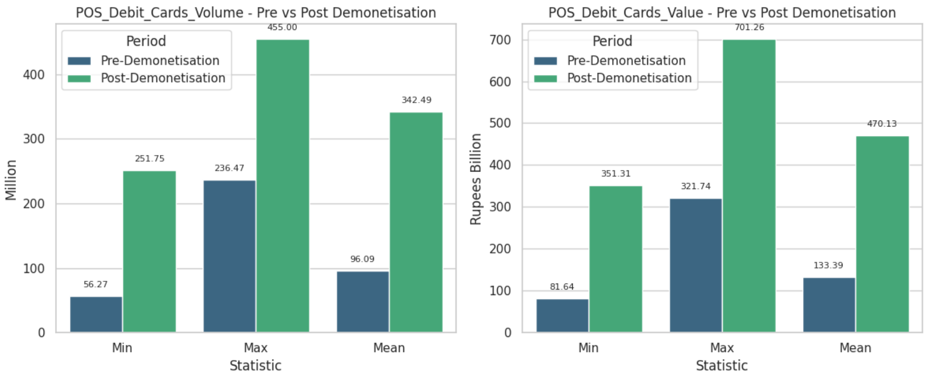

Figure 10: POS Debit Cards in pre- and post-Demonetization

The consumption of debit cards both before and after demonetization can be depicted through statistical representations in Figure 10. In the post-demonetization period, the lowest POS debit card transaction volume changed between 56.27 to 251.75 million. Whereas in the post-demonetization period, the overall POS debit card Value ranged from 140.5 to 455 million. The value of the smallest POS value debit card transaction varied from INR 81.64 to INR 351.3 and the maximum POS debit card transaction value fluctuated between INR 219 to INR 701.2. The average predicted volume of the mean is 262.9 million and the average predicted value of the mean is INR 362.5

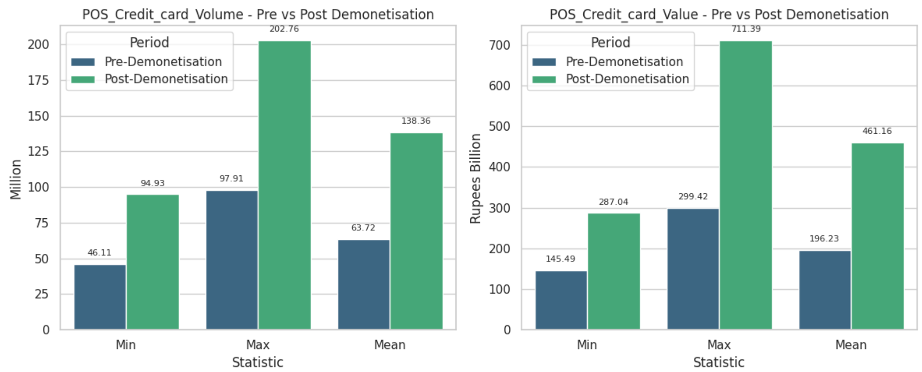

Figure 11: POS Credit Cards in pre- and post-demonetization

Figure 11 shows the usage of Credit cards at POS in terms of volume and value that happened in a 34-month window pre-demonetization and 34-month post-demonetization. In post demonetization, the minimum POS credit card volume ranged from 46.11 to 94.93 and the highest POS credit card volume varied from 88.86 to 202.76. The minimal value transaction of POS credit cards in post demonetization ranged from INR 145.49 to 287.04 billion and the highest POS credit card transaction value ranged from INR 299.4 to INR 711.39. The average predicted volume of the mean is 131.75 million and the average predicted value of the mean is INR 655.22.

In Conclusion from the above graphs, the statistical study of digital payment before and after demonetization refers to noteworthy behavior. Among all the modes, Mobile Banking is used mostly for having large transaction volume and value. Whereas POS Debit card transactions had considerably smaller value and volume, indicating a less solid outcome when compared to other modes.

4.2.4 Feature Engineering

After completion of EDA, I did the scaling of the features Min-Max Scaling is applied to the dataset. The Min-Max Scaler from the sci-kit-learn libraries was used to perform the scaling and standardized the value in a specific range of [0,1]. Feature scaling is crucial when working with deep learning models, especially with LSTM models which are sensitive to the quantity of input features.

4.2.5 Model Implementation

For the time series prediction, I have implemented one deep learning technique and another one is machine learning for comparative analysis. Firstly, I split the dataset into 80% as a training set and 20% as a test set. The deep learning method I deploy is the LSTM model using Kera's library with Tensor flow and trained for 100 epochs. Secondly, a traditional machine learning method named Exponential smoothing was implemented using the stats model library.

LSTM Model

For this study, I picked the Long Short-term Memory (LSTM) network among all other neural networks. The LSTM model is a kind of neural network that specializes in understanding patterns over a long period. RNN might have trouble with this, but LSTM is built to do so. The LSTM model is designed in such a way that learns the information and covers a longer timeframe, which makes it useful for tasks involving sequences and patterns (Albeladi, 2023) The LSTM model, I built for this study starts with the LSTM layer holding 100 units, the input shape is set to [feature=1, lookback=12]. As my problem statement is based on a regression problem, I added a Dense layer too. The model is computed using mean square error and Adam optimizer.

Exponential Smoothing Model

I deployed Exponential Smoothing (ES) as another model for time series The exponential Smoothing time series tackle provides weight to the previous dataset in an exponential decreasing order. The name exponential is given because the weight assigned to each need observation is reducing exponentially. The model gives almost the same future values as the latest past [5]. For the specified column of the training dataset, an exponential smoothing model is applied. Having a seasonal length of 12 months, it includes both additive trends and additive seasonality.

4.2.6 Model Evaluation

When examining the efficacy of the time series models that have been developed, statistical metrics such as:

Mean Squared Error (MSE): Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE were devoted. These metrics give insight into the model's accuracy and reliability in predicting the future value of past data (Yang, 2020).

Mean Absolute Error (MAE): The MAE is obtained as the average of relative differences between projected and actual values.

Root Mean Square Error (RMSE): It is determined as the square root of the mean square error. It is frequently referred to as the standard deviation of error that happens when predicting future values. This statistic is more dependent on the presence of data outliers.

---

[5] https://www.simplilearn.com/exponential-smoothing-for-time-series-forecasting-in-python-article#:~:text=The%20Exponential%20Smoothing%20time%20series,same%20as%20the%20recent%20past

$$MSE = \frac{\sum_{i=1}^{N} (O_i - T_i)^2}{N}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (O_i - T_i)^2}{N}}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |O_i - T_i|$$

Figure 12. Time Series Metric Formula (Salehi, 2021**)**

The *Oi* is defined as the actual value, *Ti* is stated as the predicted value and *N* is the number of data,

5 Implementation and Evaluation

The code for this study was developed on Google Collab using Python programming. Python is a popular programming language that data analysts use because of its flexibility and robustness. This section provides detailed implementation and evaluation of applied models respectively.

5.1 Implementation and Evaluation for Sentimental Analysis of Demonetization Tweets.

To classify the tweets into positive, negative, and neutral classifications, two deep learning models (LSTM and CNN) and one machine learning model (Naïve Bayes) were applied. The model was selected on performance criteria like accuracy, recall, precision, and F1 score. Various libraries were used in the deployment process, including pandas, NumPy, matplotlib, seaborn, TensorFlow, textblob, and Tokenizer. From the textblob library, TextBlob was used to assign polarity to emotions before building the models. Furthermore, the Tokenizer was used to transform text into a numeric format, and sequence padding was done to ensure uniform input length for useful training.

5.1.1 Experiment 1 /Evaluation of LSTM Model

For the sentimental analysis of demonetization tweets, a comprehensive approach was implemented starting from preprocessing of text. Using the Textblob package, the sentimental label was separated into binary classes, determining opinion as ‘positive’ and ‘negative’. Next, the tweet text was tokenized using the Tokenizer class. Further, the dataset was split into training (80%) and testing (20%) before applying to the LSTM model. LSTM models consist of an embedded layer used for word representation, a spatial dropout layer for normalization, and a dense layer with SoftMax activation for the binary category. The compilation of the model was done using Adam optimizer and later the model underwent training with 10 epochs. Figure 13, A classification report was generated, including detailed metrics such as precision, accuracy, and F1 score for both positive and negative sentimental classes.

Accuracy of LSTM Model: 0.9635207653045654

Classification Report of LSTM Model:

	precision	recall	f1-score	support
0	0.90	0.82	0.86	414
1	0.97	0.99	0.98	2574
accuracy			0.96	2988
macro avg	0.94	0.90	0.92	2988
weighted avg	0.96	0.96	0.96	2988

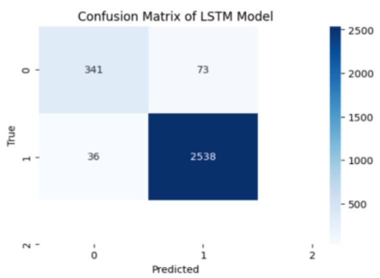

Figure 13: LSTM model results

5.1.2 Experiment 2 /Evaluation of CNN Model

A CNN model was developed in comparison with the LSTM model for sentimental analysis. The pre-processing techniques such as tokenization and padding were the same as those stated before for the LSTM model. For CNN architecture, a sequential model was constructed using keras. CNN model includes the following layers such as Embedding, permute, Conv1D, Maxpooling1D, and GlobalMaxPooling1D, which are comprised in the model, with the Dense layer having a SoftMax binary classification mechanism for activation. The CNN model was built with an error function of category cross entropy and an Adam optimizer with a learning rate of 0.01. The accuracy scores were obtained when the model was being trained. Figure 14 shows how the CNN model was evaluated in the same way that the LSTM model was, with major metrics such as precision, F1 score, and accuracy.

Accuracy of LSTM Model: 0.9635207653045654
Classification Report of LSTM Model:

	precision	recall	f1-score	support
0	0.90	0.82	0.86	414
1	0.97	0.99	0.98	2574
accuracy			0.96	2988
macro avg	0.94	0.90	0.92	2988
weighted avg	0.96	0.96	0.96	2988

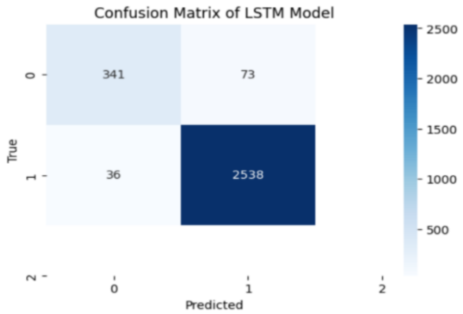

Figure 14: CNN model results

5.1.3 Experiment 2 /Evaluation of Naïve Bayes Model:

The Third model executed in this study is Naïve Bayes, particularly the Multinominal Naïve Bayes classifier from the sklearn library. The text data is altered into the vectorized form using the TFIDF vectorizer from the sklearn feature extraction library. Further, the model is split into train and test sets using the sklearn model selection library. The model is trained on the training dataset to predict the response labels for the test dataset. Next different metrics were used to analyze the model performance Figure 15.

	precision	recall	f1-score	support
0	0.88	0.72	0.79	409
1	0.96	0.98	0.97	2579
accuracy			0.95	2988
macro avg	0.92	0.85	0.88	2988
weighted avg	0.95	0.95	0.95	2988

accuracy_score: 0.9474564926372155

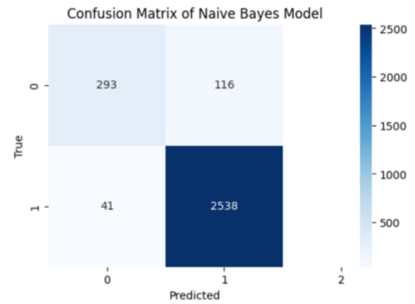

Figure 15: CNN model results

5.1.4 Sentimental Analysis Models Results Evaluation

In this section, the performance evaluation of three sentimental analysis models, LSTM, CNN, and Naïve Bayes shows an understanding of their success. Table 2 displays an exhaustive overview of its unique performance in predicting views.

Model	Classification	Precision	Recall	F1-Score	Accuracy
	Negative	0.9	0.81	0.87	
LSTM	Positive	0.97	0.98	0.98	96
	Negative	0.99	0.65	0.79	
CNN	Positive	0.95	1	0.97	95
Naïve	Negative	0.88	0.72	0.79	
Bayes	Positive	0.96	0.98	0.97	94

Table 2: Sentimental Analysis Model Evaluation

The sentimental analysis model's results show that each model has advantages and shortcomings. The LSTM model works well over both positive and negative sentiment groups with good precision (0.97 for positive attitudes and 0. 9 for negative attitudes) whereas recall states (0.98 for positive thoughts, 0.81 for negative thoughts) for demonetization. This indicates that LSTM can capture hidden sentiment patterns for demonetization tweets. On the other hand, the CNN model fails to deliver in the negative sentiment class with precision for the positive sentiment (0.95) and recall (1). This indicates CNN could have problems classifying the negative attitudes. The Naïve bays model shows lower precision and recall compared to the LSTM model. Overall, it states that the LSTM model is set apart as an exceptional performer across all metrics, displaying its effectiveness in classifying sentiment.

5.2 Implementation and Evaluation for Time series analysis on the Digital Payment dataset of India.

The time series analysis research of digital payments In India before and after demonetization uses two models: LSTM and Exponential Smoothing. These models are built using TensorFlow and stats model libraries. As it is a regression problem, the statistical metrics involved are Mean Square Error (MSE), Root means square error (RMSE), and Mean Absolute Error (MAE). The performance of all digital payments including Real Time Gross Settlement (RTGS), Retail Electronic Clearing (REC), and debit and credit card usage at POS, and mobile banking are observed and evaluated.

5.3 Experiment 1 /Evaluation of Time Series LSTM Model.

The study involves two main data frames: Transaction Volume and Transaction Value. Following steps like preprocessing, scaling, and splitting the dataset into a training set covering March 2014 to August 2018 and a test set from September 2018 to October 2019 was done for each column of the transaction volume data and transaction value data. The LSTM model designed for this study consists of one LSTM layer of 100 units with a look-back window of 12-time steps, followed by a Dense layer for prediction. The model was trained over 100 epochs with a batch size of 256. After the execution of the epoch, the prediction was made on both the test and train columns and then the column back to the original scale using inverse transformation. Furthermore, the model performances were tested using measures like MSE, RMSE, and MAE

5.3.1 Experiment 2 /Evaluation of Time Series Exponential Smoothing Model.

The Second model used in this analysis, in comparison to the LSTM model is Exponential Smoothing. Like the LSTM model, the dataset is separated into training and testing sets. To the training dataset the exponential model is applied, the model consists of additive trends, seasonality, and seasonal period of 12. Next, the model implementation was tested by comparing its prediction on the test set to the actual transaction value data.

5.3.2 Time Series Model Results Evaluation:

The table below illustrates the performance measures for both the LSTM and Exponential Smoothing model. Measures such as Mean Absolute Error (MAE), Root Mean Squared error (RMSE), and Mean Squared error (MSE) are utilized to examine the model's ability to predict values. The smaller these numbers, the better the model predicts the outcomes (S, 2012). Table 3 provides the insights of the LSTM and Exponential Smoothing (ES) models across the vital performance metrics.

Mode wise Transaction Volumn								
			Exponential_Smoothing			LSTM		
Column	Min	Max	MAE%	RMSE%	MSE%	MAE%	RMSE%	MSE%
Retail_Electronic_Clearing_Volume	105.25	839.97	10.35	10.43	1.09	5.04	3.56	0.13
Real_Time_Gross_Settlement_Volume	6.97	13.64	11.71	9.68	0.94	29.42	20.8	4.33
Mobile_Banking_Volume	10.17	1241.8	22.31	17.78	3.16	30.49	21.56	4.65
POS_Debit_Cards_Volume	56.27	455	8.93	8.83	0.78	10.87	7.69	0.59
POS_Credit_card_Volume	46.11	202.76	11.54	10.75	1.16	2.06	1.46	0.02

Mode wise Transaction Value								
			Exponential_Smoothing			LSTM		
Column	Min	Max	MAE%	RMSE%	MSE%	MAE%	RMSE%	MSE%
Real_Time_Gross_Settlement_Value	67141.64	190693.1	15.02	12.97	1.68	31.54	22.3	4.97
Retail_Electronic_Clearing_Value	4645.76	28878.55	25.94	23.32	5.44	24.63	17.42	3.03
Mobile_Banking_Value	32.96	5338.46	44.72	39.98	15.99	19.74	13.96	1.95
POS_Debit_Cards_Value	81.64	701.26	10.29	10.23	1.05	14.41	10.19	1.04
POS_Credit_card_Value	145.49	711.39	6.81	6.22	0.39	6.07	4.29	0.18

Table 3: Time series Model Evaluation

6 Discussion

For all the experiments carried out for Sentimental Analysis on Demonetization Twitter. The machine learning model like Naïve Bayes gave less precision-recall in Experiment 1, Though the Naïve Bayes model is known for its ease of use and computing efficiency, it assumes features become independent which might not be always true in the sentimental analysis problem. While the CNN model is popular for image and text classification, however, it fails to classify the negative sentiments of tweets in Experiment 2. The Long Short Term-Memory (LSTM) model was able to perform well in Experiment 3 and exhibited its ability to detect complex sentimental patterns. The overall efficiency of all three models was examined based on accuracy, precision, recall, and F1 score. Two experiments on the time series analysis of digital payment were carried out. Experiment 1 applied both exponential smoothing (ES) and Long Short Term-Memory (LSTM) models to observe value transaction data. For column such as Real-time Gross settlement volume (REC), and Mobile Banking Volume (MB), the Exponential Smoothing (ES) model performed well compared to LSTM whereas, for POS Debit Card Volume, POS Credit Card Volume, and Real-Time Gross Settlement Volume (RTGS) LSTM model gave less MAE, MSE and RMSE percentage. This states that out of five, three payment transactions volume LSTM offer good results In Experiment 2 for Real Time Gross Settlement Value (RTGS), Real time Gross settlement Value (REC), Mobile Banking Volume (MB), POS Debit Card Volume Columns Exponential Smoothing beat the LSTM and the remaining POS Credit Card Volume LSTM gave better outcome, for Experiment 2 out of five, four values of Exponential Smoothing (ES) was ahead of LSTM.

7 Conclusion and Future Work

The study aims to understand public sentiment in the month following the sudden announcement of demonetization, as well as to examine trends in digital payments after the demonetization. LSTM appeared as the best model contender in sentimental analysis employing models such as LSTM, CNN, and Naïve Bayes. This suggests positive support for the demonetization policy, as well as an interest in supporting future government initiatives. In the implementation of Time series models, experiment 1 depict that among the five payment transaction volumes, LSTM excels in three of them, and experiment 2 exponential smoothing (ES) beats the LSTM in four out of five payment transaction values. This elaborates each model works differently for different payment transactions. Although having different values, both models exhibit lower MSE, RMSE, and MAE showing their efficacy in capturing and evaluating transaction trends. This shows a consistent increase in digital payment in India both in terms of value and volume. The finding points to a positive trend in the country's adoption of digital payments over time.

Extending the sentimental analysis to include the Hindi language is an interesting direction for future development. Since most Indians can read and write the Hindi language, this addition could provide a more improved description of the popular mood. Including statistics connected to the Government of India's lately implemented new payment systems data[6]. More insights can be obtained into appearing trends in digital payments by applying advanced forecasting methods like FB Prophet, Bi-LSTM, and so on.

8 References

Alam, K. N., 2023. Retracted: Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. *Hindawi Computational and Mathematical Methods in Medicine,* Volume 2023, p. 16.

Albeladi, K., 2023. Time Series Forecasting using LSTM and ARIMA. *(IJACSA) International Journal of Advanced Computer Science and Applications,* 14(1), p. 8.

Asati, A., n.d. A Comparative Study On Forecasting Consumer Price Index Of India Amongst XGBoost, Theta, ARIMA, Prophet And LSTM Algorithms.. *PGD-Machine Learning and Artificial Intelligence,* p. 8.

Bansal, N., 2018. Progress in Digital Banking After Demonetization: Some Evidence. *The IUP Journal of Bank Management,* Volume 17, p. 11.

Bansal, N., 2018. Progress in Digital Banking After Demonetization: Some Evidence. *The IUP Journal of Bank Management,* 17(2), p. 11.

Chamsukhee, V., n.d. Time Series Forecasting of E-Databases, Subscription Mahidol University Library with Exponential Smoothing , LSTM, and ARIMA Models. p. 6.

Chavali, K., 2019. Demonetisation and its Impact on Digitisation in India. *Managment Dynamics,* 19(1), p. 17.

Chopra, R., 2017. Impact of Demonetization on Indian Economy. *Global Journal of Enterprise Information System,* 9(2), p. 4.

[6] https://dbie.rbi.org.in/#/dbie/statistics/Payment%20Systems

Dar, K., 2017. Demonetization: Sentiment and Retweet Analysis. *Proceedings of the International Conference on Inventive Computing and Informatics,* p. 6.

Darliansyah, A., 2019. Long-Term Trends in Public Sentiment in Indian Demonetisation Policy. *ResearchGate,* p. 12.

Dr Jai Kishan Karahyla, M., 2023. Predicting Price Direction of Bitcoin based on Hybrid Model of LSTM and Dense Neural Network Approach. *Proceedings of the Fourth International Conference on Electronics and Sustainable Communication Systems* , p. 6.

Goularas, D., 2019. Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data. *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications* , p. 6.

Goyal, V., 2019. Sentiment Analysis of English-Punjabi Code Mixed Social Media Content for Agriculture Domain. *International Conference on Information Systems and Computer Networks,* p. 6.

Hindocha, M., 2019. The Impact of Demonetization on Electronic Fund Transfers. *The IUP Journal of Bank Management,* 18(2), p. 13.

Huang, F., 2022. Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis. *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS,* 33(9), p. 14.

Khatik, S. K., 2018. A Study on Pre- and Post-Analysis of Demonetization Period: Issues and Challenges. *International Journal of Business Analytics and Intelligence,* 6(2), p. 10.

KOLYA, A. K., 2020. Sentiment classification with GST tweet data on LSTM based on polarity-popularity model. *Indian Academy of Sciences,* p. 17.

Kumar, A., 2019. Demonetization in India: Good or Bad in Context of Social Media. *INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING AND SOFTWARE ENGINEERING,* p. 6.

Manocha, S., 2019. The impact of demonetization on digital payment transactions: a statistical study. *International Conference on Advancements in Computing & Management,* p. 7.

Nithin, M., 2019. Has Demonetisation Pushed Digitalisation in India? Some Counter Evidences. *Journal of Business Thought,* Volume 9, p. 13.

Ohlan, M., 2019. Demonetization of Indian Economy : A Review on the Effects and Reactions. *Indian Journal of Extension Education,* 55(3), p. 8.

Pirani, M., 2022. A Comparative Analysis of ARIMA, GRU, LSTM and BiLSTM on Financial Time Series Forecasting. *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics,* p.6.

Pramod, D., 2023. Exchange Rate Prediction with Machine Learning, Deep Learning, and Time Series Methods Using Alternative Data. *International Conference on Advancement in Computation & Computer Technologies,*p.6.

Salehi, F., 2021. Influence of Infrared Drying on Drying Kinetics of Apple Slices Coated with Basil Seed and Xanthan. *International Journal of Fruit Science,* 21(1), p. 10.

Siami-Namini, S., 2018. A Comparison of ARIMA and LSTM in Forecasting Time Series. *IEEE International Conference on Machine Learning and Applications,* p. 8.

Singh, P., 2017. Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. *The Korean Institute of Communications and Information Sciences (KICS),* p. 6.

Srinivas, A. C. M. V., 2020. Sentiment Analysis using Neural Network and LSTM. *IOP Conf. Series: Materials Science and Engineering ,* p. 8.

S, S., 2012. PERFORMANCE COMPARISON OF TIME SERIES DATA USING PREDICTIVE DATA MINING. *Advances in Information Mining ,* 4(1), p. 10.

Unnithan, K. S., 2022. Analyzing Tweet Data to Identify the Impact of Demonetarization. *Proceedings of International Conference on Emerging Technologies and Intelligent Systems,* p. 7.

Vempaty, L. N., 2016. ANALYSIS OF TWITTER DATA BY EXTRACTING THE TWEETS ON DEMONETIZATION. *International Journal of Data Analytics (IJDA),* 01(01), p. 6.

Yang, H., 2020. Short-term carbon emissions forecast for aviation industry in Shanghai. *Journal of Cleaner Production 275,* p. 12.