National College of Ireland

Sentiment Analysis on Demonetization and rise of Digital Payments using Deep Learning: India

MSc Research Project
MSc Data Analytics

Nishant Bharti
Student ID: x21148686

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Nishant Bharti
Student ID:	x21148686
Programme:	MSc Data Analytics
Year:	2018
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	20/12/2018
Project Title:	Sentiment Analysis on Demonetization and rise of Digital Payments using Deep Learning: India
Word Count:	XXX
Page Count:	20

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:	
Date:	30th January 2023

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):	

Sentiment Analysis on Demonetization and rise of Digital Payments using Deep Learning: India

Nishant Bharti
x21148686

Abstract

Due to digital disruption, digital payments have gained appeal in last few years. As e-commerce and digital marketplaces increase, electronic payment processing has become a financial innovation priority. In 2016, the Indian government proclaimed and implemented demonetization. Demonetization is analysed using sentiment and exploratory data. Building a neural network deep learning framework is difficult. This research analyses the top three digital payment systems using a neural network long-short term model and exploratory data. To study public opinion on demonetization, a transitional period leading to the spread of Paytm, PhonePe, and Google Pay till 2021. Satisfactory model accuracy was 82%. Both sentiment analysis and deep learning on payments app data accomplished the goal. The importance of online digital payment systems was shown using python libraries. With the success of this research, machine learning and deep learning converged on sentiment analysis, resulting in the growth of online payment systems in a few years.

1 Introduction

When a currency is no longer in use, it ceases to be recognized as legal tender and has no value. In an effort to eradicate illicit money and curb smuggling, several nations have opted to reinstate defunct currencies as legal tender, a process known as demonetization.

1.1 Motivation and background of Demonetization

India before 2016 was facing a heavy shadow of black money which laid to corruption, cross-border smuggling. On November 8, 2016, India demonetized 500 and 1,000 rupee banknotes. New 500 and 2,000 banknotes were also revealed. The Indian PM said the decision would diminish the shadow economy, enhance cashless transactions, and cut crime and terrorism financing with illegal and counterfeit money. India is likely to go to digital payments. It happened while smuggling, black money, and corruption peaked in India. At Initial stage, it was very difficult to get new currency notes handy, getting new notes standing in queues for hour that to limited amount can be exchanged was spreading anger amongst people. The aims of the demonetization process were said to have been moved about by the government, which was regarded as shifting the goalposts. As it became apparent that almost all the cash was being exchanged, the goals were expanded to include making India a cashless economy, neutralization of money held by Maoists, terrorists, and human traffickers, amongst other things. The initial stated goal was to curb black money, corruption, and terrorism.

It is said that one decision demonetization served various purposes and solved problems of the economy. There was no substantial growth of tax collection before demonetization but after two years of it made tremendous increase till 2017. In 2021, over 700 million Indians use debit cards and 30% use credit cards. Demonetization caused instability, and people began adopting digital payment systems. Demonetization helped people understand 'Cashless India'. Now more than ever, it's important for people to have access to formal financial services, hence cashless transactions are gaining popularity. The widespread adoption of digital payment systems is gaining innovation and bolstering the case for a fully digitalized economy. As more and more companies move their operations online, the country's payment infrastructure went a complete overhaul to accommodate the growing digital economy. In the 2015–2016 fiscal year, the currency-to-GDP ratio stood at 12.1%. As a result of demonetisation, it went down to 8.8% in the 2016–17 fiscal year, but it went back up to 10.9% in the 2017–18 fiscal year. It took three years for online digital payments to enter the market, but by the time the global pandemic of 2019–2020 brought the country to a standstill, online payment services like Paytm, UPI, Phonepe, and Googlepay had established themselves as the dominant payment options. The sudden surge in online activity was a major factor in the closure of the system. The whole financial sector shifted to an online payment system.

1.2 Research Question

When the top two greatest denominations of currency in circulation (500 and 1000 rupees) were suddenly and unexpectedly discarded without any additional warning or backup plan to substitute new denominations in banks, people panicked. For deep learning to be useful for the digital payment system, it is necessary to first understand why sentiment analysis on demonetization is being conducted.

RQ: *What extent can demonetization using sentiment analysis technique enhance digital wallets and reduce cash payments to support India cashless economy idea?*

It is essential to grasp the public's perspective and feelings about the demonetization imposed on the country and the future goals of the cashless economy.

SQ: *Can Deep Learning model enhance the rise of digital wallets to support Indian Government initiative of cashless economy?*

Deep learning is applied separately to online digital payments. If we get positive feedback and see an increase, we know that people also like cashless platforms. Thus, sentiment analysis and deep learning are two different things. However, when combined both, this research aims to show that sentiment analysis on demonetization both idea are in favour and have positive feedback works as a bridge to the rise in digital wallets and online payment systems.

1.3 Research Objectives and Contributions

Following research objectives were derived to address research question.

Obj.1 A critical review of sentiment analysis on demonetization and traditional existing payment methods from Obj.2 to Obj.5. The results of the reviewed literature helped in understanding different types of sentiment analysis techniques and deep learning models on digital payments system from Obj.2 to Obj.5 .

Obj.2 Sentiment Analysis on Demonetization.
(a) Broadly investigate the trends of tweets and retweets to study sentiments of people on demonetization.
(b) Derive the behaviour of the tweets and sentiment type that is, most frequent words and extract positive and negative distribution of tweets.
(c) Experiment using Multinomial Naïve Bayes classifier model.

Obj.3 Exploratory Data Analysis on online digital payments.
(a) Apply Exploratory EDA on top three online digital payments platforms emerged after demonetization and obtain best platform based on reviews.
(b) Rise of all three platforms Phonepay, Google,pay and Paytm for three consecutive years 2019-2021.

Obj.4 Implementation of Deep learning on Phonepay, Google,pay and Paytm using LSTM model.
(a) Comparison of top three online digital payment platforms.
(b) Apply Deep Learning model using LSTM model on all three payments apps data.

Obj.5 Evaluation and results presented.
(a) Sentiment analysis
(b) Evaluation of exploratory data analysis
(c) Evaluation of deep learning model

1.4 Conclusion

The rest of the technical report was addressed as follows: Chapter 2 shows related work on sentiment analysis within demonetization and a critical review of deep learning on digital wallets. The modified Demonetization methodology approach was also explained in Chapter 3. The goals from section 1.3 are put into action, evaluated, and the results are shown in chapter 4. Last, chapter 5 talks about the research question, and chapter 6 is the conclusion of demonstrated results and recommend future work.

2 A Review on Sentiment Analysis on Demonetization and Digital Wallets using Deep Learning

2.1 Intorduction

The Indian government's decision to demonetize currency was the most significant step toward improving the country's economy that it had ever taken. At the same time as black money and corruption were at record highs, cross-border smuggling was increased on daily basis. The issuance of new 500 and 1000 rupee notes by the government had

a considerable influence on the behavior of those who were holding illicit money. When the government first suggested "Cashless India," those who had illegal funds had their full costs brought to light. The United States has gotten more acclimated to making use of online payment tools as a result of the government's introduction of the UPI and Phone Pay applications for online purchases. Because of this, the issue of illegal money has been significantly mitigated. Because online payment applications have become such an important part of the Indian economy, a substantial amount of scientific research is required in order to gain insights into people's reactions to demonetization and to investigate the rise of online payment applications in India. Specifically, this research is required to investigate the rise of online payment applications in India. Since the announcement of demonetization in 2016 and the subsequent push to transform India into a "cashless" nation, several studies have been carried out to investigate people's feelings on the topic and the ways in which online payment resources might be adapted.

2.2 Impact of Sentiment Mining practices on Demonetization

Opinion mining is another name for sentiment analysis. The goal of this method is to ascertain the author's opinion about the offered service or item.Singh et al. (2018) noted that many backed demonetization by tweeting many positive messages on November 8, 2016. Ravinder Singh and Kahlon reviewed nation-wide and state-wide geolocation analyses to explain people's discontent. They separated it into two stages and found a negative decline, with 4,551 (32.42%) positive tweets and 4,675 (33.32%) negative tweets. In the second phase, favorable tweets made up barely 1% of the 11,294 gathered. Similarly, Arun et al. (2017) analyzed Twitter sentiment using data cleansing, bigrams, polarity, sentiment scores, and graphical approaches.

Rasika Wagh and Payal Wagh and Punde (2018) developed two sentiment analysis approaches: supervised and unsupervised learning. They favoured supervised machine learning algorithms like Nave-Bayes, SVM, and Maximum-Entropy. Another paper, Mohbey et al. (2022) analyzed and compared sentiment analysis algorithms based on polarity. Sentiment analysis on the Twitter data set using machine learning algorithms was the focus of Dhanya and Harish (2018). They thought about the first month of data to analyze. We examine the outcomes of analyzing the same collection of 5000 tweets using three different machine learning methods naive bayes classifier, support vector machine and decision tree. We compared several classifiers and found that the Naive Bayes classifier was the most accurate. Niharika Kumar Kumar (2017) similarly studied political, natural catastrophe, sports, and entertainment data six months after demonetization. The author notes that people's views were divided on the issue and claims that news coverage has a major impact on how the general public feels about it.Maindola et al. (2018) examines the impact of alternative monetary systems on social network sentiment analysis using IBM Watson. After a year of demonetization, the authors analysed tweets and other sites where individuals communicated their thoughts on the wallets. The report included some fascinating details regarding the state of digital payment alternatives in India and the attitudes of its users. After time went people understood the motive behind such impactful sudden flip in cash to online digital payment system. People started thinking in positive side of demonetization and future goals. Roy et al. (2017) emphasized the optimistic slant, noting that over half of all tweets were found to be in favor of this action by the PM of India. Almost a quarter (22%) of all tweets were classified as just informative. While 33% of tweets were found to have unfavourable feelings, most of them

were directed towards the way the act was carried out rather than the act itself.

T Srinivas et al. (2019a) conducted Hierarchical clustering, Support Vector Machine, Nave Bayes, and Logistic Regression and then provided a graph-based visual model for classification. These methodologies' results were compared and analyzed on factors such algorithm accuracy, prediction, and confusion matrix generation. Logistic regression's 94% accuracy beats the other method. Tweepy is a Python library used for accessing Twitter data . It is also called as a wrapper class for accessing twitter. Tweepy was presented in Vaid et al. (2017), which analyzed whether or not the general public saw demonetization as a beneficial development for the nation. The author employed the classify emotion(_) and classify polarity(_) R functions, each of which is cited as a key tool for sentiment analysis in R. The author said the research could have included more tweets and a nation with 1.25 billion inhabitants for better results. T Srinivas et al. (2019b) gave a graphic model to help readers better comprehend classification after discussing hierarchical clustering, Logistic Regression, Naive Bayes, and Support Vector Machine. Several conclusions are derived from a comparison and analysis of these methods based on criteria such as algorithmic correctness, prediction accuracy, and confusion matrix generation. Using geolocation data and a lexicon-based methodology, the authors of Ray et al. (2018) analyzed user sentiment on Twitter in the five weeks leading up to the policy announcement.

Kumar and Singh (2019) presented public opinion on money laundering, demonetization, and the Sensex and Nifty. According to the research, pro-demonetization advocates cited benefits such as the elimination of terror funding and the transition of the real estate market to digital payments. Author compared TextBlob and python library of Natural Language Toolkit (NLTK) and favoured NLTK over TextBlob. It stated that most of the tweets were posted from the metro cities of India. Black money was the most discussed topic by the people on Twitter during demonetization.

2.3 Review on Impact of Demonetization

Higher denominations being eliminated will have numerous negative effects. Similar phrases was discussed Sharma (2017) effects on parallel economy, money supply, demand, pricing, economic entities, GDP, internet transactions, and alternative ways of payment. Without appropriate and effective planning, the demonetization-driven cash shortage has left India's informal sector, which accounts for 40% of GDP, unviable. Many who paid in cash are having server issues. Exchange lines consume hours of working time. Frail and elderly people have a hard time getting fresh notes. On the other hand Aswani (n.d.),clearly stated that people have negative opinions on demonetization. Smartphone users who are familiar with web technologies may simply use digital payments. SHILPA and AMULYA (n.d.) said that youngsters use digital channels for online purchases more than middle-aged and older individuals. Thus, digital modes must be made stronger, safer, and more versatile so individuals of all ages purchase online. Digital marketing reaches clients of all ages, and its effect on youth and society creates a good platform for digital marketers.

2.4 A critical review on Deep learning and Machine Learning techniques

It's only been around two to three years since cashless payment systems like digital payment and wallets become the normal. Indeed, the massive diversion occurred after the Covid-19 pandemic. Basically, the idea is that there aren't that many articles and not a lot of data to use machine learning algorithms on. The following articles elaborated on the various methods and applications of machine learning in the context of demonetization.

Ananthajothi et al. (2022) used the machine learning technique Neural Network (NN) and the Self Adaptive Beetle Swarm Optimization (SA-BSO) method to choose optimal features. SA-BSO was used to optimize RNN's hidden neurons and feature selection. Author discarded using existing approaches because of problems with mistake rate, sarcasm, multi-polarity, and word ambiguity. Low accuracy and neutral feelings aren't minimized. To tackle these issues, the SA-BSO method is developed to maximize the accuracy of demonetization tweet sentiment categorization. And concluded with SA-BSO-RNN model is accurate and reduces error. Latent Dirichlet Allocation (LDA), a topic modelling approach, was proven on tweets in Harshvardhan et al. (2021). The model evaluates words and themes using measures like saliency and relevance, and it calculates distances between subjects using multidimensional scaling. The model gave some interesting results that it can perform even better on a bigger corpus. Kannan et al. (2019) research paper uses a support vector machine model to anticipate demonetization data (PAD-SVM) Semantic analysis is also used to determine user sentiment levels and the compound polarity of each tweet. Predictive analysis focuses on how individuals think about issues and how society responds to them. Gupta and Singal (2017), demonstrated sentiment analysis used filter tweets and compare accuracy and efficiency at same time. The author implemented Naive Bayes and SVM model and obtained 49% and 51% accuracy

2.5 Conclusion, existing challenges and gaps

Despite the fact that several research on demonetization utilizing sentiment analysis have been conducted, most of them have been conducted with a very small number of data, especially considering that India still has a population of 1.25 billion even after five years have passed. However, there have been far fewer research conducted on this problem in the areas of machine learning and deep learning.To the best of the candidate's knowledge, there hasn't been nearly enough research on digital payment employing aspect-based sentiment analysis. The vast majority of scholars have examined customer evaluations using polarity score or sentiment analysis. However, the efficacy of machine learning and deep learning for this application is still up for debate. There isn't enough time to gather the necessary data to obtain a full picture of demonetization in only two or three months. The quantity of data required to get precise consumer insights is large.

Hence objective Obj.1 discussed in Section 1.3 have fully achieved successfully.

3 Methodology and Design Specification

3.1 Introduction

Tweets and retweets are sometimes generated and distributed with the express purpose of acquiring trust or spreading fake news. During demonetization vary initial stage that is at end of the year 2016, these tweets and rumours had the potential to sway the actual logic and purpose of the campaign and it is best to have a system to make it real time.

3.2 Modified Methodology Approach Used

Cross-industry method referred to as CRISP-DM and this was applied in this study because all the six stages below individually contribute and fulfil from start to end requirements of the objective mentioned in this project. All the perspective are covered thoroughly at each stages of the adapted CRISP-DM methodology. Below is the figure of adapted stages and working of each stage of the adapted CRISP-DM methodology.

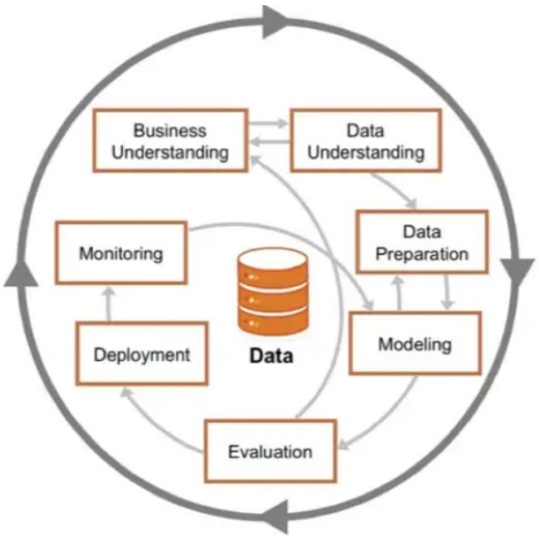

Figure 1: Adapted CrispDM Methodology Steps

Despite the fact that the CRISP-DM approach consists of six phases, these stages are adaptable, which makes it possible to go back and forth between the various levels. In figure 1, in the first phase, define research questions and goals with the indepth business understanding, which requires an in-depth understanding of the desired outcomes of your work. During this stage, the cost of the work is calculated and the right tools for the job are chosen. The next stage, which is data interpretation. Initial data collection is summarized, investigated, and checked for quality here. The following step is the data preparation phase, which consists of five activities (select, clean, build, and format) applied to the final data sets for modelling. The fourth step, modelling, comes right after the data preparation phase. There are four steps involved in data modelling: deciding which modelling approach to use, creating test data, developing the model, and providing access to the model.

3.3 The process of Data Extraction and Architectural Design

After first stage of the adapted methodology and understand business needs, a data collection on which the research is to be done. In this study, data is captured and retrieved from open-source website Kaggle. Follwoing Figure 2 shows the entire components of the implementation in the project.

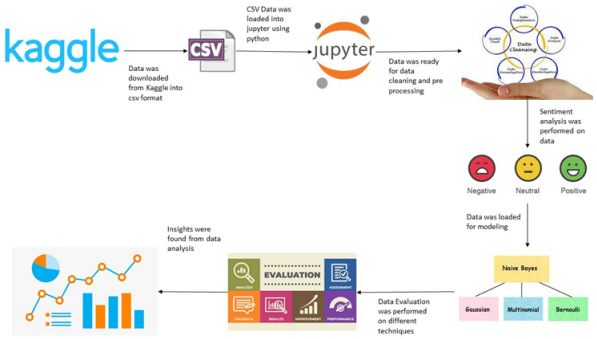

Figure 2: Overview of data Preparing and modelling (Ref: Self Draw.IO)

Opting Crisp-DM methodology and its stages combined with design architecture of the this research illustrates the overall proposed start to end working of sentiment analysis. The Kaggle data was downloaded as a.csv file and then imported into a jupyter notebook for additional analysis. In the next step, data cleaning is performed to eliminate duplicates and get the data suitable for modelling. After the data has been cleaned, a sentiment analysis model is applied to it, and a thorough grasp of the public's feelings towards demonetization is gained. Information evaluation involves deciphering the model's output in order to draw conclusions and then representing those conclusions graphically.

3.4 Working of Sentiment Analysis and Deep Learning Model

The research mainly works around two approaches that is Sentiment Analysis and Deep Learning Framework LSTM model. This section in the same two sub-section are as follows:

i) **Sentiment Analysis** : It is a subfield of text classification that falls under the umbrella of Natural Language Processing (NLP). Sentiment analysis, in its simplest form, is labelling a text with an emotional category, such as "happy," "sad," "neutral," etc.

ii) **Deep Learning LSTM method**: Recurrent neural networks (RNNs) are a kind of artificial neural network that can learn long-term relationships, which is notably useful in sequence prediction issues. Unlike methods that can only handle single data points, such as photos, LSTM can analyse the full data sequence because to its feedback links. This is useful in a variety of contexts, including voice recognition and automatic translation. Short-Term Memory Recurrent Neural Networks (LSTM) are a subclass of RNNs that have shown to perform very well across many different types of challenges. The below architecture shows the working of the model:

In Figure 3 below, the line that connects one node's output to the inputs of other nodes is shown as carrying a whole vector in the figure that can be seen above. While the yellow boxes are learnt neural network layers, the pink circles represent pointwise operations such as vector addition. Concatenation is shown by lines merging together, but a line crossing

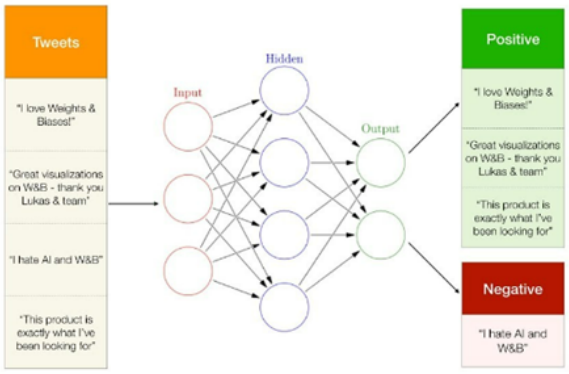

Figure 3: Working and Flow Diagram of Sentiment Analysis[1]

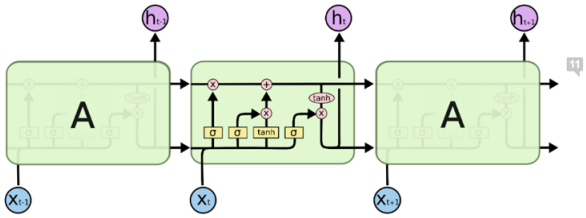

Figure 4: LSTM Model Architecture

indicates that its content is being duplicated and the copies are traveling to separate places.

iii) **Exploratory Data Analysis**: The term "exploratory data analysis" is often used to describe the crucial process of doing preliminary investigations on data in order to find patterns, to notice anomalies, to test hypotheses, and to evaluate assumptions with the use of summary statistics and graphical representations.

Figure 5: Exploratory Data Analysis Steps (Ref:Self, Draw.io)

Getting the raw data suitable for application of the model required some work in the form of extraction and cleaning. Figure 5 is a flowchart depicting the fundamental analytical concept. In this study, we used EDA to analyze data from three different mobile payment services, namely Phonepay, Googlepay, and Paytm. Lastly, the output of EDA process is the cleaned data on which Deep Learning model was applied.

3.5 Conclusion

As a result, this section devoted its whole to discussing all of the components, techniques, and approach in great detail. In this part, each of the three stages of this study—that is, sentiment analysis, EDA, and the deep learning framework—along with the elements of each that are most appropriate for this research were applied and explained. This project's phases and requirements were defined, in large part, by the use of an adapted version of the CRISP-DM methodology. Before applying the model or the principles, every piece of data that was utilized in the study was cleansed so that we could receive accurate findings. At last, the working and idea behind both sentiment analysis and LSTM model was thoroughly elaborated.

4 Sentiment Analysis Implementation, Evaluation and Results of Deep Learning Demonetization Models

4.1 Introduction

In this section, the findings of testing the model's efficiency will be presented. It was determined to carry out the tests on the datasets. To determine how well the proposed model and workings in demonstrating sentiment analysis and deep learning model, in real time scenarios. This study aims to analyse the opinions of people on demonetization play role of bridge to rise digital payments using deep learning.

4.2 Software languages, libraries and tools used

Python- According to the Python website, Python is an interpreted, object-oriented programming language with dynamic semantics. Its high-level in-built data models, together with Dynamic Typing and Dynamic Binding which allow for rapid developing applications. Python's straightforward, easy-to-learn syntax also contributes to the language's readability. Python above features and easy to interpret and understand stand out Among all other languages.
Libraries- Since Python is so agile, numerous coders have used it to make their own machine learning libraries. Researchers in machine learning use Python more and more due to the abundance of libraries, like TensorFlow, Scikit-Learn, NumPy, Keras, PyTorch, etc. Some of the libraries that were used for this project are:
NLTK (Natural Language Toolkit)- It is used for large portion of the information at your access is unstructured and may be read by humans. This data has to be preprocessed before it is analysed using a computer program. This library works the concept of tokenizing were it allows to work on smaller pieces of text that are relatively coherent and meaningful even outside the context of rest of the text. Since, in this research first part, data is in form of unstructured tweets, tokenizing by word or sentence make more feasible to use NLTK.
VADER (Valence Aware Dictionary for Sentiment Reasoning)- VADER's sentiment analysis is based is a lexicon and rule-based sentiment analysis tool that is specifically adapt to sentiments expressed in social media. The working is combination of a sentiment lexicon is a collection of words or other lexical elements that have been assigned positive or negative labels based on their underlying semantic orientation. Thus, it is played a vital role in deriving sentiment analysis in this case.

Sklearn- The Multinomial Naïve Bayes using sklearn library have a discrete features of word count in text classification. It is used in data pre-processing part and feature extraction stages.
Matplotlib and Seaborn- Python's Matplotlib module is a graphics library that may be used to create charts and graphs from data. Incorporating Pandas and NumPy into the software is straightforward. Seaborn simplifies working with Pandas' data frames. The Matplotlib extensions offers a more user-friendly collection of tools for making stunning images in Python.
Keras- Keras is a Python-based deep learning application programming interface that operates on top of TensorFlow, a machine learning framework. Rapid experimentation was a primary design goal throughout development.

4.3 Aspect Based Sentiment Analysis on Demonetization

4.3.1 Implementation

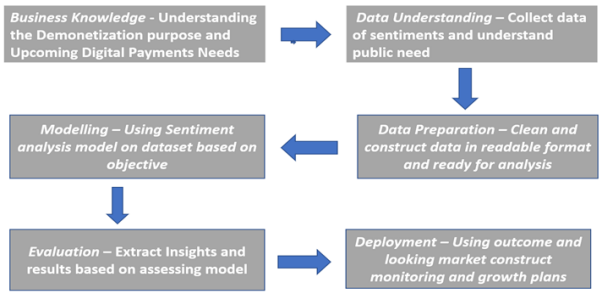

Figure 6: Demonetization Steps and Flow Diagram (Ref: Self/ Draw.io)

The process flow diagram for sentiment analysis of demonetization shown in (Fig.6). Process starts from understanding the deep knowledge and logic behind demonetization. Data pre-processing occurs after data collection and understanding. Data pre-processing involves data cleaning operations such as removing null values, stop words, and white spaces, among others. Once the data is clean, Python code is run on it to get the sentiment score from the reviews. The algorithm was evaluated after receiving the complete data set with attributes and sentiment score. Finally, visualization is done with python matplotlib library.

Since two data frames are used in this paper that is, first to study sentiment on tweets and second data is combined three data sets of UPI payments apps into one data frame and then deep learning model was applied, thus data pre-processing is divided in two sections to make it more transparent and understandable.

4.3.2 Data Pre-processing for Sentiment Analysis

The data consists of more than 14000 tweets which was enough to carry out the reliable insights from the research. In data cleaning, it was found there were some redundancies in data like stop words, retweets and blank spaces which detected and discarded from the data. Hence result 10441 tweets were left after cleaning data.

```
Total Number of tweets
10441

Number of POSITIVE tweets -  6370
Percentage of POSITIVE tweets
61.00948185039747

Number of NEGATIVE tweets -  4071
Percentage of NEGATIVE tweets
38.99051814960253
```

Figure 7: Results after Data Cleaning

In above figure 7, we can derive the data was cleaned and their type of data with percentage contribution.

4.3.3 Evaluation and results of Sentiment Analysis

This subsection contains the results of the model's effectiveness tests. The decision to run the tests on the datasets was made. Aiming to evaluate the efficacy of the suggested model and its implementation, done on sentiment analysis.

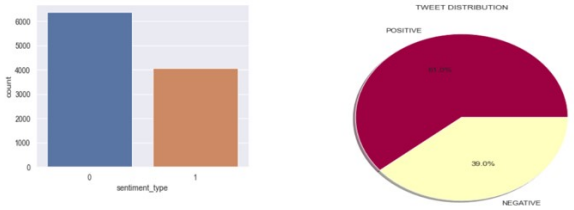

Figure 8: Sentiment type and distribution

From figure.8 and 9, it was determined, during the first phase of the research project on opinions, that most of the tweets belonged to the positive feedback. This indicates that most people had a favourable perspective towards the demonetization of the currency. it is apparent that more than 6,000 tweets, or 61% of the total number of tweets, support the demonetization. And as can be seen in figure.9, the vast majority of individuals are content with the choice to do the demonetization,

4.3.3 Evaluation and results of Multinomial Naive Bayes model

Multinomial Naive Bayes assumes the existence of a feature vector in which each element corresponds to the frequency with which the feature occurs in the data (or, very often, its frequency) .

In below figure 10, its represntes the accuracy of the model with 95.87% training accuracy and 92.05% testing accuracy and the algorithm was able to get accuracy of the model, precision, recall, F1-score is 95.87%, 93%, 94%, 93% respectively, with confusion matrix shown below in figure 11.

Objective Obj.2 (a), 2(b), 2(c) and objective 5(a) has been achieved on sentiment analysis which partially answers the research question mentioned in the section 1.2.

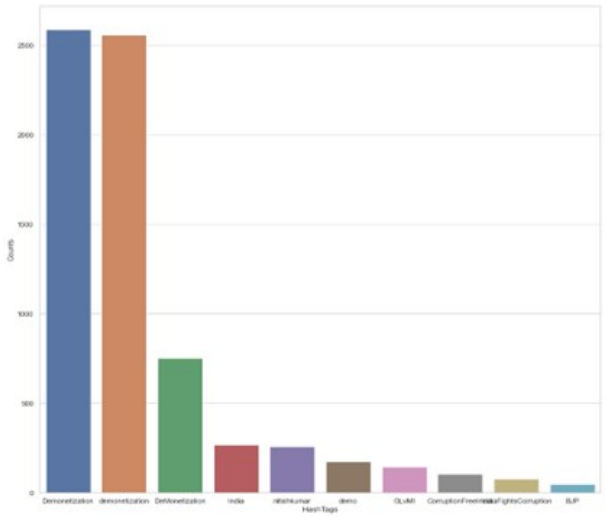

Figure 9: Sentiment type and distribution

```
Multinomial Naive Bayes
MultinomialNB()

Training Accuracy Score - 95.87  %
Testing Accuracy Score - 92.05  %

 Following is the Classification Report for Multinomial Naive Bayes -
              precision    recall  f1-score   support

           0       0.93      0.94      0.93      1895
           1       0.91      0.88      0.90      1238

    accuracy                           0.92      3133
   macro avg       0.92      0.91      0.92      3133
weighted avg       0.92      0.92      0.92      3133


Testing AUC Score -  91.41  %
Training  AUC Score -  95.63  %
```

Figure 10: Mutlinomial Naive Bayes Results

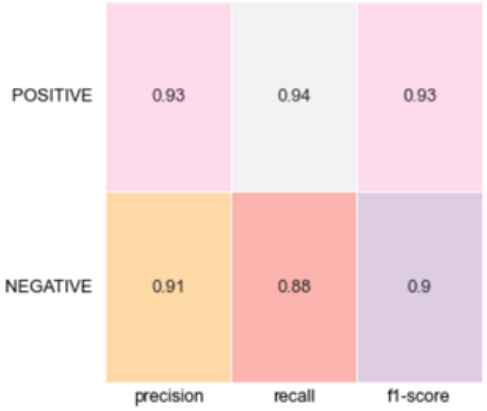

Figure 11: Mutlinomial Naive Bayes Results:Part 2

4.4 Exploratory Data Analysis Results on Online Digital Payments

4.4.1 Implementation of Exploratory Data Analysis

Exploratory Data Analysis is a component of both data exploration and data preparation. During this stage, the raw data is put through an analysis process in which it is all cleaned up and feature extraction is performed. After cleaning data all three datasets were combine for further evaluation and get insights from it. The results of this process are then used as the basis for modeling and visualization, which produce reliable results.

4.4.2 Data Pre-Processing

In data cleaning step, outliers, stop words, null values were detected which was handled using python. All three digital payment apps data, googlepay, phonepay and paytm data were investigated separately as shown in below figure 12, with the output of all three payment apps.

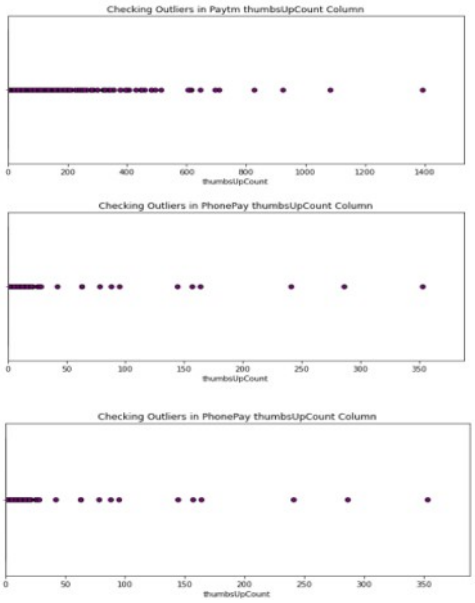

Figure 12: Outliers in Googlepay, Phonepay and Paytm Data

On further investigation, it was found that all three data detected outliers and were handled and removed from the data. The columns which are not in use were dropped.

4.4.3 Evaluation and Results

Exploratory data analysis of the three fastest-growing methods of digital payment in 2019-2021 based on evaluations provides the key insights and findings shown in the following charts from the second stage of the study.

In figure 13 below it is evident from the chart that, initially, people did not favour using online payment systems; nevertheless, there was a rapid spike in the number of people using these systems in the year 2021, with paytm and phonpe being utilized at the largest level.

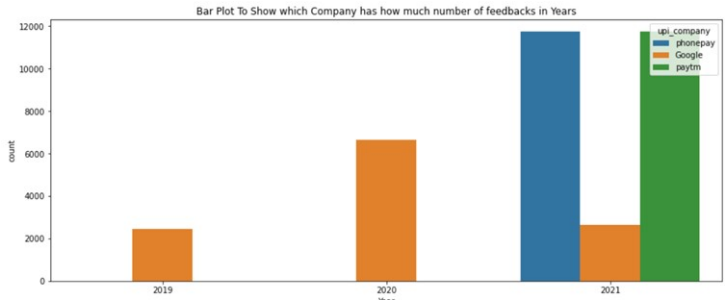

Figure 13: EDA reuslts: Part 1

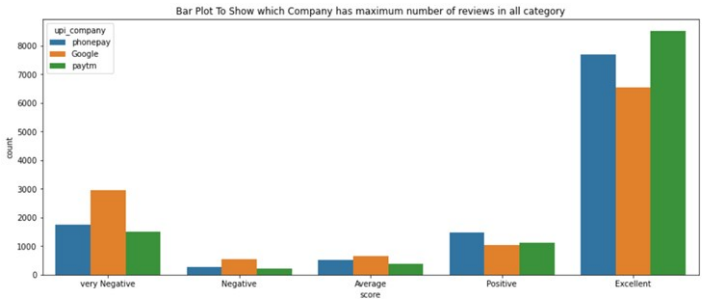

Figure 14: EDA reuslts: Part 2

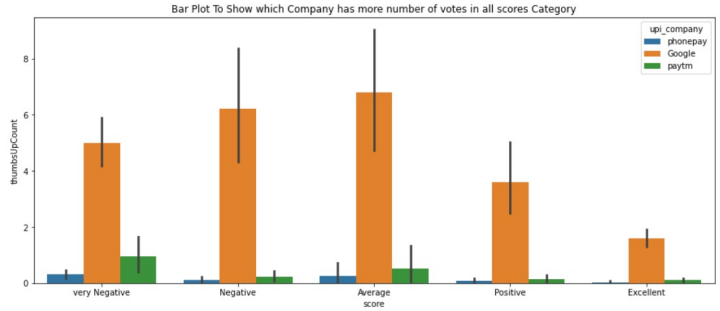

Figure 15: EDA reuslts: Part 3

The data shown in the preceding figure 14 and figure 15 makes it abundantly evident that users choose Paytm as the most successful online payment platform in the year 2021, and Phonepe as the second most preferred application for making payments. From figure 15, it was clear that googlepay is ahead in terms of most votes from excellent to very negative

Therefore, objective Obj.3(a), Obj.3(b) and Obj.5(a), has been achieved and the data results of this section is used in next section that is to implement deep learning model.

4.5 Implementation of Deep learning on Phonepay, Google,pay and Paytm using LSTM model.

4.5.1 Introduction

After conducting exploratory data analysis on the data collected by combine apps, as shown in the flowchart located above, a deep learning LSTM model is built. Then, with the assistance of tokenizer logic that is associated with the model, text classification is carried out word by word throughout all of the responses. Owing to the fact that the binary cross-entropy function from the keras library is utilized in order to improve text categorization. In its simplest form, it determines the loss in cross-entropy that occurs between the real labels and the anticipated labels. Below figure 16 shows the flow diagram of LSTM model.

Figure 16: Flow Diagram of LSTM model

4.5.2 Steps in Implementing LSTM model

Step1 1: Use the Python packages Keras and nltk to import the necessary dependencies and write the activation functions.
Step 2: Initializing the biases and weight matrices.
Step 3: Multiplying forget gate with last cell state to forget irrelevant tokens.
Step 4: Sigmoid Activation decides which values to take in and than transforms new tokens to vectors.
Step 5: Calculate the present cell state.
Step 6: Calculate the output state.

4.5.1 Evaluation and results of LSTM model

As discussed in 4.4.1 section, Before applying LSTM model, deep review system was studied to get proper insights on top three platform.

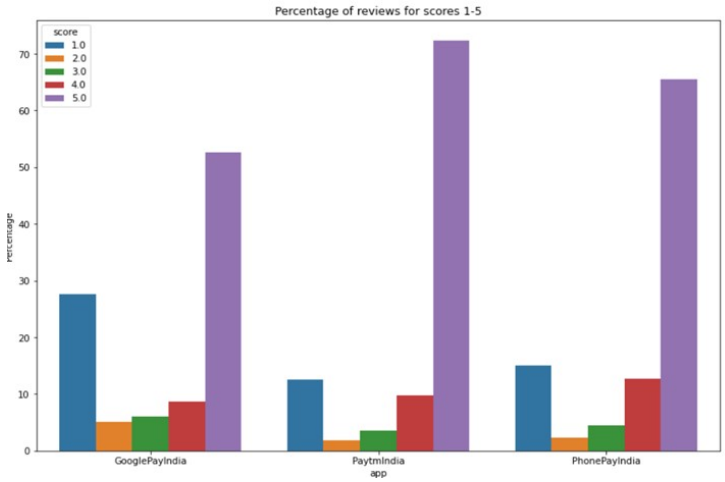

Figure 17: Result of LSTM model on top three apps

It was obvious, based on the data provided below figure 16, that Paytm stand out from the other two online payment applications over three consecutive years by receiving a review rating of five out of five stars whereas Phonepay was the second most favourable online digital payment app.

After successfully applying deep learning framework LSTM model, the model achieved accuracy of 82% overall as shown in figure 17.

Thus, objective Obj.4(a), Obj.4(b) and Objective Obj 5.(c), has been successfully achieved and also fulfil the second part of research question(Ref subsection 1.2).

4.6 Conclusion

Thus outputs presented in this section and implementation and evaluation of the techniques, model output shown in figure 18 below and analysis combined together enables us to answer the research question and sub-research question (Ref section 1.2) have been achieved and the results can help us understand that sentiment on demonetization laid to sudden rise of digital payments in India and was on the road to achieve 'cahsless economy'

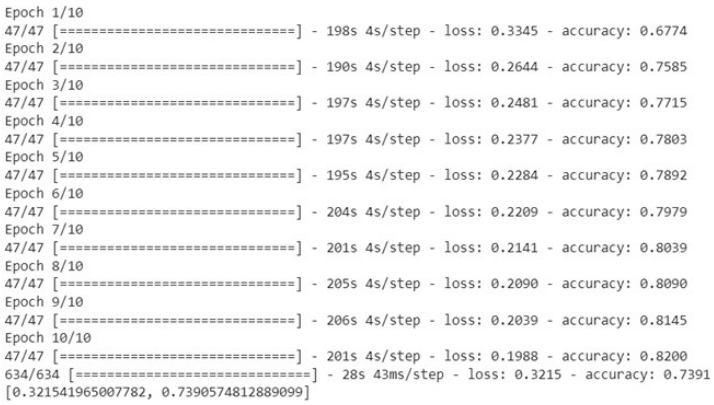

Figure 18: LSTM Model Implementation

5 Discussion

To answer the research question: *How can we apply sentiment analysis on demonetization in India and analyse the rise of digital wallets and online payment system to achieve idea of 'cashlessIndia' using deep learning?*
The project focused on sentiment analysis and deep learning combined together on demonetization and its impact on exiting cash system converting into totally opposite digital payment system.To achieve this, first study phase revealed that most people support the demonetization initiative and are open to the concept of a cashless economy (discussed in section 4.1). Exploratory data analysis and a deep learning model were used to determine which cashless, secure, and transparent payments applications will develop in the three years after demonetization (2019-2021). Paytm and Phonepe were determined to have the highest level of support (82%) among these apps(discussed in section 4.4). Hence, decision of Demonetization had positive impact which later it played important role to expand digital payments to fulfil 'cashless economy' idea.

6 Conclusion and Future Work

Sentiment analysis on demonetization was carried out in a research study employing several natural language processing packages in Python. Data has been tested the deep learning model. Long short-term memory (LSTM) model obtained excellent and adequate accuracy in deep learning. This study is separated into three sections: gathering internet reviews and doing aspect level sentiment analysis, exploratory data analysis, and developing a deep learning model. All three stages of this project were interdependent, with people's sentiments playing a significant part in how the deep learning phase fulfilled the project's aim.

To the best of the candidate's knowledge, insufficient study into the application of deep learning to the field of digital wallets has been conducted during the previous few of years. This study helps to fill blanks and bridge the gaps. This study makes a contribution to the digital economy of India and to the transition of other financial sectors to an online system entirely. According to what is mentioned in (section 1.2), all of the goals of the research question have been met.

Future Work: Only the English language was taken into consideration for the purpose of this study. Since every language has its own unique structure and set of grammatical rules, this discovery opens up a significant amount of doors for potential future work in other languages. While deploying the deep learning model was the sole method that produced excellent accuracy, this aspect of the problem might be investigated in further work, and even more accuracy could be attained using alternative models that were deployed in this domain. In addition, the study has been carried out during a time frame of three months. It is possible that hundreds of gigabytes of data might be gathered in a very short amount of time, and that similar study could be carried out in order to get more precise insights. And as a last point, it is possible that in future work, supervised machine learning methods as well as unsupervised machine learning approaches will be used.

Acknowledgements I would like to thank my supervisor, Dr. Catherine Mulwa, in particular for all the help and advice she gave me all semester.I would like to thank my parents for being there for me and believing in me. I also want to thank my best friends Pooja Rakate, Renuka Sarukar, Aayusha Kashilkar, and Shruti Desai, who helped me when things were hard.

References

Ananthajothi, K., Karthikayani, K. and Prabha, R. (2022). Explicit and implicit oriented aspect-based sentiment analysis with optimal feature selection and deep learning for demonetization in india, *Data & Knowledge Engineering* **142**: 102092.

Arun, K., Srinagesh, A. and Ramesh, M. (2017). Twitter sentiment analysis on demonetization tweets in india using r language, *International Journal of Computer Engineering In Research Trends* **4**(6): 252–258.

Aswani, C. P. M. (n.d.). Impact analysis of demonetization using r.

Dhanya, N. and Harish, U. (2018). Sentiment analysis of twitter data on demonetization using machine learning techniques, *Computational vision and bio inspired computing*, Springer, pp. 227–237.

Gupta, F. and Singal, S. (2017). Sentiment analysis of the demonitization of economy 2016 india, regionwise, *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, IEEE, pp. 693–696.

Harshvardhan, G., Gourisaria, M. K., Sahu, A., Rautaray, S. S. and Pandey, M. (2021). Topic modelling twitterati sentiments using latent dirichlet allocation during demonetization, *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, pp. 811–815.

Kannan, N., Sivasubramanian, S., Kaliappan, M., Vimal, S. and Suresh, A. (2019). Predictive big data analytic on demonetization data using support vector machine, *Cluster Computing* **22**(6): 14709–14720.

Kumar, A. and Singh, J. P. (2019). Demonetization in india: Good or bad in context of social media, *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*.

Kumar, N. (2017). Sentiment analysis of twitter messages: Demonetization a use case, *2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, IEEE, pp. 1–5.

Maindola, P., Singhal, N. and Dubey, A. D. (2018). Sentiment analysis of digital wallets and upi systems in india post demonetization using ibm watson, *2018 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, pp. 1–6.

Mohbey, K. K., Bakariya, B. and Kalal, V. (2022). A study and comparison of sentiment analysis techniques using demonetization: case study, *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines*, IGI Global, pp. 1441–1454.

Ray, P., Chakrabarti, A., Ganguli, B. and Das, P. K. (2018). Demonetization and its aftermath: an analysis based on twitter sentiments, *Sādhanā* **43**(11): 1–10.

Roy, K., Kohli, D., Kumar, R. K. S., Sahgal, R. and Yu, W.-B. (2017). Sentiment analysis of twitter data for demonetization in india: a text mining approach, *Issues in Information Systems* **18**(4): 9–15.

Sharma, M. (2017). Twitter sentiment analysis on demonetization an initiative government of india, *International Journal of Recent Trends in Engineering & Research, 3 (486–490)* .

SHILPA, N. and AMULYA, M. (n.d.). Impact of demonetization on e-payments in india, *CHRIST COLLEGE* p. 1.

Singh, P., Sawhney, R. S. and Kahlon, K. S. (2018). Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by indian government, *ICT Express* **4**(3): 124–129.

T Srinivas, A. S., Govinda, K., Ramasubbareddy, S. and Swetha, E. (2019a). Sentimental analysis of demonetization over twitter data using machine learning, *Journal of Computational and Theoretical Nanoscience* **16**(5-6): 2055–2058.

T Srinivas, A. S., Govinda, K., Ramasubbareddy, S. and Swetha, E. (2019b). Sentimental analysis of demonetization over twitter data using machine learning, *Journal of Computational and Theoretical Nanoscience* **16**(5-6): 2055–2058.

Vaid, I., Kumar, P., Rawat, S. and Choudhury, T. (2017). Sentiment analysis of indian currency demonetization, *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, IEEE, pp. 85–90.

Wagh, R. and Punde, P. (2018). Survey on sentiment analysis using twitter dataset, *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, pp. 208–211.