Sentiment Analysis and Evolution of Cashless India: Pre-Covid, During Covid and Post Covid.

MSc Research Project
Data Analytics

Vaishnavi Gaikwad

Student ID: x23109203

School of Computing
National College of Ireland

Supervisor: Abubakar Siddig

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ....VAISHNAVI GAIKWAD.......................................

Student ID: ....X23109203..

Programme: ....M.Sc. Data Analytics...... **Year:** 2023-2024

Module: Research Project

Supervisor: ... Abubakar Siddig ....

Submission Due Date ....11/08/2024..

Project Title: **Sentiment Analysis and Evolution of Cashless India: Pre-Covid, During Covid and Post Covid** ....

Word Count: ......6000............... **Page Count**.........15............

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: ....VAISHNAVI GAIKWAD......................................

Date: ....11/08/2024..

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):	

Sentiment Analysis and Evolution of Cashless India: Pre-Covid, During Covid and Post Covid.

I. *Abstract*

The research considers the sentiments of the users and performs sentiment analysis to understand their response to the evolution of cashless payment, through Google Pay, Phone Pe, and Paytm UPI apps.. The data for the same was collected by web scraping app reviews, and were then categorised as positive, negative and neutral, in the periods namely categorised as Pre-Covid, Post-Covid, and during Covid periods. Later, the study used machine learning and deep learning models, viz: LSTM (long short-term memory), ARIMA and SARIMA, to predict future transaction volumes of these apps based on empirical UPI data from the NPCI website.
The study pre-processed transaction volume data from 2017 to 2024 and forecasted usage from 2025 onwards to understand the trajectory of cashless payments in India. The models were considerably successful in predicting the transaction volumes for Google Pay, Phone Pe, and Paytm, identifying which app is will most likely be used in the future. The study provides valuable insights for industry practitioners and academic researchers into the evolution of cashless payment apps in India, wherein the models can help in providing better insights for better financial forecasting.

The findings can have a significant impact on fintech companies, for using data-driven strategies and for optimal operational efficiencies that support the global adoption and evolution of secure cashless payment systems. The effectiveness and advantages of LSTM, ARIMA, and SARIMA models is also highlighted in our research which will thereby help in forecasting UPI transaction over time, contributing to the field of financial time series analysis.

***Keywords**:* LSTM Neural Networks, Financial Time Series Forecasting, UPI Transaction Volumes, Google Pay, Paytm, Phone Pe, Sentiment Analysis, Pre-Covid, Post-Covid, During-Covid.

II. *Introduction*

The sudden surge in the digital market industry in India has significantly contributed to the evolution of cashless payment transactions, as the country heads in the direction of a digital transformation in the payments domain. Cashless payments began gaining popularity in the year 2016, yet the widespread adoption was limited because of the large landscape and the population. However, from 2016 onwards, the landscape changed dramatically with the introduction of the Unified Payment Interface (UPI), which revolutionized digital payments by enabling instant money transfers between interpersonal bank accounts through mobile apps. The study finds out that Google Pay, Phone Pe, and Paytm have been the most widely used apps in a market crowded with lot of other UPI apps. This high usage of mentioned app is in terms of transaction volume and user adoption.[1]
The study analyses app reviews during the Pre-Covid, Post-Covid, and During Covid periods. These reviews were categorized into positive, negative, and neutral sentiments. The analysis highlights that during the pre-Covid period, cashless payment apps were not widely used by maximum population. [2] However, during the Covid period, usage surged due to concerns over virus transmission through physical currency. In the post-Covid period, public sentiment towards these apps became increasingly positive due to their convenience and probably because the common public was used to the apps by then. The research shows which apps provided the best service and received the most positive feedback across all three periods, demonstrating the evolution of Google Pay, Phone Pe, and Paytm.[3]

The study also made use of the financial data from the National Payments Corporation of India (NPCI) from the years 2017 to 2024. Using the empirical transaction volume data, we applied machine learning and deep learning models, viz: LSTM, ARIMA, and SARIMA, to predict the future growth of cashless payment apps.[4] This research tries to verify the performance of these models, evaluates their prediction accuracy and then discusses its academic and real-world implications. Furthermore, the findings offer valuable insights for stakeholders, helping them to make informed decisions and improve services, while also contributing to the broader field in the financial forecasting and strategies in digital payment.[5]

Research Questions:

1)How is the Sentiment Analysis performed concerning the Mobile Wallets i.e. Phone Pe, Google Pay, and Paytm by scraping reviews from app (Apple App Store)?

This question defines how the sentiments of the common public are gathered and analyzed such as their reviews about the mobile wallet applications i.e. Google Pay, Phone Pe, and Paytm from mobile app which will help us to get the sentiments based on "positive", "negative" or "neutral".

2)What were the effects of the pandemic COVID-19 on the growth and usage patterns of mobile applications i.e. phone pe, google pay, and Paytm. Which have served as mobile wallets since its inception, analyzing trends before, during, and after the pandemic using LSTM, ARIMA and SARIMA model.

In these questions the effects of covid 19 on mobile wallets are determined since its inception from the era before after and during covid using LSTM, ARIMA and SARIMA model, to find its trends.

III. *Literature Review:*

Researchers and scientists have used varied methods for sentiment analysis on web apps through scraping. This section focuses on previous work on sentiment analysis using web scraping and its applications in our research. [Daniel Glez-Peña and Anália Lourenço's][6] research on web scraping in the biomedical domain underlines the importance of web services for data integration and exchange, according to which web scraping emerged as a critical method to bridge gaps left behind by web services, enabling efficient data extraction. In our study, we decided to implement web scraping on App store and Google Play Store applications.

[Swathi Venkatakrishnan et al][7]. verified the sentiment analysis through mobile app analytics, focusing on customer behaviour through app metrics such as reviews, ratings, and operating functions. The study deployed deep learning and machine learning techniques using the KERAS library, while the research utilizes the NLTK library for sentiment analysis. [Emitza Guzman and Walid Maalej][8] showcased the effectiveness of fine-grained sentiment analysis, particularly for user feedback in app reviews. They highlighted the challenges of manual review analysis and the advantages of automated sentiment analysis using Natural Language Processing (NLP) and machine learning. The approach thus offers scalable solutions, and in the research, we applied the same to UPI applications in India.[Ashish Joy, G. Rejikumar, and M. Dhanya][9] conducted sentiment analysis on UPI applications in India, noticing the fast growth of e-payments following UPI's introduction. Their research underlined the importance of understanding user perceptions regarding app security, usability, and overall satisfaction. By analysing user sentiments, service providers could improve app features, enhance security, and streamline interfaces, contributing to the larger picture of making India digital and giving impetus to a cashless economy.[Mahesh A. and Ganesh Bhat][10] analysed the growth of UPI in India by examining transaction volumes, which was similar to the approach of forecasting UPI app growth using machine learning and deep learning models. The study provided valuable insights into the expansion of UPI, which we applied in our analysis of Google Pay, Phone Pe, and Paytm. [Dr. Asha Pachpande and colleagues'][11] focus was to forecast the growth of UPI applications using NPCI data, which motivated the research to aim for accurate growth predictions for UPI apps.[R. Khatwani et al.][12] observed growth and penetration of digital payments in India, offering valuable insights for the research on UPI's impact in India. Finally, [Nancy Goel et al][13]. conducted exploratory research on India's digital payments landscape using machine learning techniques, guiding the research's prediction and analysis of the growth of UPI apps in India.

The study of previous work on this topic underlines the importance of sentiment analysis and machine learning techniques in understanding and predicting the evolution of UPI applications in India. The studies verify our approach to analysing user sentiments and forecasting the growth of cashless payment apps in India. Taking cues from thew previous works on the topic, our work aims to contribute to both academic knowledge and practical strategies for enhancing digital payment systems, which would finally support the broader picture of financial inclusion and a cashless economy in India.

Summary of the Literature Review with Research Questions.

The literature emphasizes using, sentiment analysis through web scraping and machine learning to understand user feedback on mobile wallets like Phone Pe, Google Pay, and Paytm. Sentiment analysis of app reviews is automated using NLP to classify sentiments as positive, negative, or neutral. Additionally, models like LSTM, ARIMA, and SARIMA are applied to analyse the growth of UPI apps, with a focus on the COVID-19 pandemic's impact on digital payment trends.

Using research questions

1. Sentiment Analysis: How can user reviews from app stores help analyse the sentiments toward mobile wallets like Phone Pe, Google Pay, and Paytm?
2. Pandemic Impact: How did the COVID-19 pandemic influence the growth and usage of mobile wallets, and what trends can model like LSTM, ARIMA, and SARIMA reveal?

IV. *Research Methodology:*

Data Collection:

In this research methodology, initially let’s discuss about the steps employed in this research to scrape the reviews from Apple app store for sentiment analysis to understand the usage and growth of the cashless payment applications during Pre-Covid, Post-Covid and during covid in India. The reviews are scraped from Apple App store to get the public reviews of cashless payment applications in India that is phone pe, Google Pay, Paytm. The reviews are scraped from year 2017 to 2024.

Tools and Libraries:

Programming Languages: Python

Libraries: ‘app_store_scraper’, ‘pandas’

Data Storage: CSV files

Steps for Scraping reviews from Appl App Store:

- The python library ‘app_store_scraper’ used to scrape reviews from apple app store, also the panda’s library is used for data storage and data manipulation.
- To scrape reviews from the applications like Google Pay phone pay and Paytm we created an instance of class App Store in which we need to pass the parameters of application such as the ID of application on App Store with app name and the country from which the app belongs to
  Ex: For Paytm, country='in', app_name='Paytm-secure-upi-payments', app_id='473941634'.
- Once the instance is created, we specified the year from which the reviews should be extracted, for every application we extracted the reviews from 2017 to 2024. Which is divided into pre post and during COVID period.
- Once the reviews are extracted from every application, we created a separate CSV file of all the applications right from 2017 to 2024.

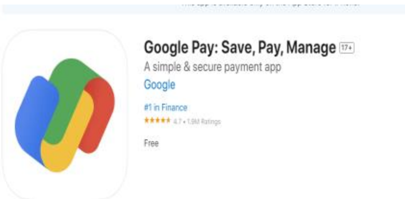


Fig 1: Google Pay Logo

Fig 2: Paytm App Logo

Fig3: Phone Pe App logo

Data Pre-processing:

In Data Pre-Processing, Once the app scraping is done for every year from 2017 to 2024 Now, we have the data sets for Google Pay reviews from 2017 to 2024, Phone Pe Reviews from 2017-204 similarly we have for Paytm reviews from 2017-204.

To perform sentiment analysis on the scraped reviews from app, let us look at the columns present in all three data sets of Google Pay reviews, Phone Pe Reviews and Paytm reviews. Data Sets named as: 'GooglePaySentiments.csv', 'PhonePeSentiments.csv', 'PaytmSentiments.csv'.

Columns: below are the columns present in reviews dataset

```
date,review,rating,isEdited,title,userName,developerResponse
```

Fig4: Columns from Sentiment Data sets

Let us look at the basic information of every dataset to understand size of data set, number of columns, number of rows:

```
Google Pay Data Info:
Size: 17661
Number of Rows: 2523
Number of Columns: 7
```

```
PhonePe Data Info:
Size: 9846
Number of Rows: 1641
Number of Columns: 6
```

```
Paytm Data Info:
Size: 12000
Number of Rows: 2000
Number of Columns: 6
```

Fig 5: Sentiments dataset information

Also in data pre-processing, we dropped the columns that are no longer necessary to perform sentiment analysis, columns that are dropped 'is Edited', 'developer Response' if exist.

To Understand how data is structured and understand basic Statistics from Data Set we have performed EDA process.

Exploratory Data Analysis (EDA) Process: To perform EDA process for every data set, we tried to find the null values in every Datasets statistic values and do understand the values in data set, some of the values of every data set are printed

EDA for Google Pay reviews: Helps us to get performance metrics for google pay data.

```
EDA for Google Pay:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2523 entries, 0 to 2522
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   date      2523 non-null   object
 1   review    2523 non-null   object
 2   rating    2523 non-null   int64
 3   title     2523 non-null   object
 4   userName  2523 non-null   object
dtypes: int64(1), object(4)
memory usage: 98.7+ KB
```

```
            rating
count  2523.000000
mean      3.054300
std       1.670944
min       1.000000
25%       1.000000
50%       3.000000
75%       5.000000
max       5.000000
```

```
               date                                              review  \
0  22-11-2017 17:06  Hi,\n\nFirst of all, i did not expect such a P...
1  10-11-2017 05:25  InShort: Your own money may get stuck for 60 d...
2  03-10-2017 17:14  I totally would recommend this app for all the...
3  18-11-2017 14:58  Tez is not reliable. The money does not get cr...
4  22-09-2017 19:20  The app is much slower than BHIM, PhonePe and ...

   rating                                           title         userName
0       1                                    Pathetic app  rottenblade397
1       1  Highly Buggy and unreliable app, save your tim...          Sajalb
2       5                                 Simple and great      ashu gupta
3       5  Money does not get credited in the receiver's ...       audittime
4       5                             Good but can improve      Diveshp747
```

Fig 6: EDA for Google Pay Sentiments:

EDA for Phone Pe defines statistics and performance metrics for Phone pe.

```
EDA for PhonePe:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1641 entries, 0 to 1640
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   date      1641 non-null   object
 1   review    1641 non-null   object
 2   rating    1641 non-null   int64
 3   userName  1641 non-null   object
 4   title     1641 non-null   object
```

```
            rating
count  1641.000000
mean      3.790372
std       1.651291
min       1.000000
25%       2.000000
50%       5.000000
75%       5.000000
max       5.000000
```

```
               date                                             review  \
0  03-10-2017 15:53  There are very few apps that one would search ...
1  11-11-2017 09:39  I had a problem with postpaid bill payment and...
2  12-11-2017 17:47  They might promise you cashbacks but never ful...
3  29-09-2017 01:47  In my iphone 6 this app is working but transac...
4  22-10-2017 11:27  I have used many payment apps like Paytm, airt...

   rating      userName                           title
0       5   Amit Prabhu                     Awesome app
1       5  Ashik dharba           Poor customer service
2       1        Veer24             Far Below standards
3       5     Jesvin007  App not working for more than a month
4       5   lovieesingh                Easy and very fast
```

Fig 7: EDA for Phone Pe Sentiments

EDA for Paytm reviews defines statistics and performance metrics for Paytm.

```
EDA for Paytm:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   date      2000 non-null   object
 1   review    2000 non-null   object
 2   rating    2000 non-null   int64
 3   title     2000 non-null   object
 4   userName  2000 non-null   object
```

```
            rating
count  2000.000000
mean      3.153500
std       1.803764
min       1.000000
25%       1.000000
50%       4.000000
75%       5.000000
max       5.000000
```

```
               date                                             review  \
0  04-10-2017 15:47  Ultimate app ..Easy to use Paytm is very very ...
1  22-11-2017 17:11  Paytm has been a trusted payments app all thes...
2  20-10-2017 09:40  Guys, please don't book bus tickets through pa...
3  17-10-2017 04:01  I just got addicted to paytm, cant even think ...
4  11-11-2017 15:06  Although the app does what it's supposed to do...

   rating                                          title         userName
0       5  PAYTM APP IS GOD GIFT FOR HUMAN Beings....👍👍...          Vik1582
1       4                           Trusted Payments App          iamalap
2       1                   Worst service in the universe  Sreeprakash C J
3       5                             Addicted to paytm          Diza@123
4       5                         Works good - feels bad      iamAnkitRoy
```

Fig 8: EDA for Paytm reviews

Sentiment Analysis:

To perform sentiment analysis, we used NLTK libraries, VADER lexicon is a sentiment analysis tool which is used to analyse text, it categorizes text into positive, negative and neutral sentiments. Initially downloaded VADER sentiment analyser, converted 'date' to a date time. In function to calculate sentiment score using VADER, applied sentiment analysis to each dataset to analyse text the 'review' column is used, based on review column the text is analysed and by analysing sentiments the sentiment score is calculated, now let us understand how the data is distributed in Pre-Covid, During-Covid, Post-Covid. According to official government website in India the Pre-Covid period is considered as starting from '2017-01-01' TO '2019-01-01', During-Covid period start '2020-01-01', During Covid END '2021-12-31' and Post Covid period is defined as starting from '2022-01-01' to '2024-01-01'. Based on sentiment score calculated for Pre-Covid, Post-Covid and During-Covid. Once the sentiment score is calculated of each period, let us classify the sentiment based on score to understand the positive negative and neutral sentiments for Pre-Covid, Post-Covid and During Covid. To classify sentiments according to sentiment score, if sentiment score >0.05 the sentiments determine the positive reviews about Google Pay, Phone pe and Paytm . if the score < -0.05 it determines the negative sentiments of all three applications and the sentiment with remaining score determines the neutral sentiments.

After classifying and analysing the sentiments for Pre-Covid, Post-Covid and during covid based on the public reviews, let us see the google pay sentiments, phone pe sentiments, and Paytm sentiments along with its sentiment score which gives us sentiments into positive negative or neutral

```
Google Pay Sentiments:
{'Google Pay': {'PRE COVID': {'Positive': 348, 'Negative': 202, 'Neutral': 21}, 'DURING COVID': {'Positive': 492, 'Negative': 3
53, 'Neutral': 41}, 'POST COVID': {'Positive': 579, 'Negative': 403, 'Neutral': 81}}}
```

Fig 9: Analysed Sentiments for Google Pay Reviews

As we can see above, for google pay in the period of Pre-Covid there are more positive sentiments of public which defines that the google pay were most used and liked by every common person in India, During Covid also the Google Pay has more positive sentiments by people ,while in the Post Covid period to the Google pay performed really very well in the market and made common public life easy to use cashless payment device.

Now let us see for Phone Pe and Paytm.

```
PhonePe Sentiments:
{'PhonePe': {'PRE COVID': {'Positive': 372, 'Negative': 233, 'Neutral': 43}, 'DURING COVID': {'Positive': 197, 'Negative': 128,
'Neutral': 33}, 'POST COVID': {'Positive': 372, 'Negative': 211, 'Neutral': 51}}}
```

Fig 10: Analysed Sentiments for Google Pay Reviews

As seen above Phone pe show's us the positive and negative sentiments during each period

```
Paytm Sentiments:
{'Paytm': {'PRE COVID': {'Negative': 501, 'Positive': 465, 'Neutral': 47}, 'DURING COVID': {'Negative': 245, 'Positive': 226,
'Neutral': 16}, 'POST COVID': {'Negative': 245, 'Positive': 233, 'Neutral': 20}}}
```

Fig 11: Analysed Sentiments for Google Pay Reviews

Whereas the Paytm sentiments shows us the negative reviews by the people.

Along with the dataset of app scraping to analyse sentiments. In order to understand the growth of Cashless payment applications from pre-covid, during covid and post covid, also to predict the cashless payment application that is to be used in future. We are using the NPCI finance dataset (https://www.npci.org.in/what-we-do/upi/upi-ecosystem-statistics#innerTabTwoJun24).

NPCI is the Indian government website that operates retail payments and settlements in India

of UPI, the cashless payment applications run based on UPI (Unified Payment Interface) that operates payments between peer-to-peer and person-to merchant transactions. The extracted data from NPCI is right from 2022 to 2024 that is available on website currently to get the data for particular periods that is for Pre-Covid, Post-Covid, During-Covid the remaining data is synthetically generated and verified from 2017 to 2021 based on the NPCI government website. Let us see the columns in the finance data from NPCI and the synthetic data that is generated based on NPCI finance UPI data.

```
 #   Column
---  ------
 0   Date
 1   UPI Banks
 2   Volume (Mn) By Customers
 3   Value (Cr) by Customers
 4   Month
 5   Year
```

Fig12: Columns in NPCI dataset

Synthetic Data is generated based on previous data with columns:

- Date: Date Column represents the date on which the value and volume of UPI is calculated.
- UPI Banks: This column represents whether the Cashless Payment is done by GOGGLE PAY, PHONE PE or PAYTM.
- Volume By Customers (Mn): This column represents the number of customers using UPI transaction per day.
- Value by Customers (Cr): This column represents the set value by customers per UPI Transaction.
- Month: Represents the month of UPI transaction.
- Year: Represents the year of UPI transaction.

The Data is generated from APRIL 2017- APRIL 2024 and extracting the Data from APRIL 2022 -APRIL 2024 from NPCI, The Data sets is merged and named as Finance Data. Now. let us understand the date and pre-processing step for finance data.

```
Finance Data Data Info:
Size: 46548
Number of Rows: 7758
Number of Columns: 6
```

Fig13: NPCI Finance data information

To understand the statistics from the finance data we have performed the EDA process which gives us result's as:

```
EDA for Finance Data:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7758 entries, 0 to 7757
Data columns (total 6 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Date                      7758 non-null   object
 1   UPI Banks                 7758 non-null   object
 2   Volume (Mn) By Customers  7758 non-null   float64
 3   Value (Cr) by Customers   7758 non-null   float64
 4   Month                     7758 non-null   object
 5   Year                      7758 non-null   int64
```

```
       Volume (Mn) By Customers  Value (Cr) by Customers
count               7758.000000             7.758000e+03
mean                1083.214275             1.738832e+05
std                 1549.393278             2.471517e+05
min                    0.147684             5.083874e+02
25%                    9.373419             6.038547e+03
50%                  280.794767             3.744521e+04
75%                 1488.640926             2.762632e+05
max                 7714.489131             1.246154e+06
```

```
         Date   UPI Banks  Volume (Mn) By Customers  Value (Cr) by Customers  \
0  01-04-2017  Google Pay                  1.678545              3065.692692
1  02-04-2017  Google Pay                  1.423209              3110.892611
2  03-04-2017  Google Pay                  1.609248              3080.447128
3  04-04-2017  Google Pay                  1.588376              3089.287146
4  05-04-2017  Google Pay                  1.537275              3090.666494

   Month  Year
0  April  2017
1  April  2017
2  April  2017
3  April  2017
```

Fig14: Finance data EDA statistics

Now the finance dataset is ready for PRE-COVID, POST-COVID and DURING-COVID for UPI Banks.

Let us see the correlation between finance data and app reviews data to understand how the reviews are correlated with the transaction volume of customers.

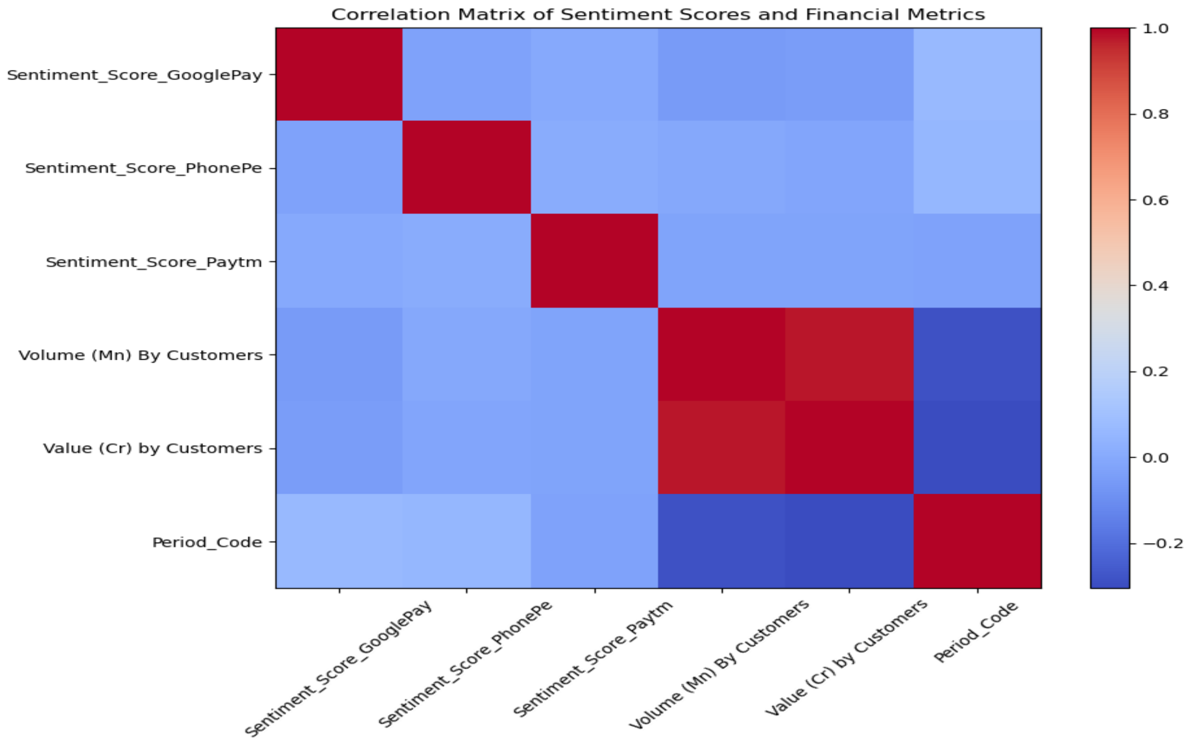


Fig15: Correlation Matrix of Review Data and Finance (NPCI) data

As the sentiments are analyzed and the sentiment scores are calculated for each app the correlation matrix tells us that as the sentiments represents public opinion which are indirectly correlated to finance data, the finance data gives us the accurate calculation of transaction volume to predict the growth of cashless payment app evolution.

To predict the actual values on past data, predicted values on past data and future forecasting for evolution of cashless payment apps we used finance data.

Data Integration:

In data integration of finance data, the date is converted into datetime format. Once cashless payment app reviews are scraped, and sentiment analysis is performed. Then the finance data is used to determine the growth of cashless payment applications in future by using Machine Learning models and the prediction of most used Cashless payment applications in future is determined in further analysis.

To understand the growth of cashless payment in India for Phone Pe, Google Pay, Paytm by using Time-series Machine learning modules we are trying to determine growth of cashless payment application since its inception and predict the application that will be more used in future.

1]Long Short-term Memory Algorithm:

STEP 1: Imported libraries in python are MinMaxScaler, mean_squared_error, Sequential, LSTM, Dense, Dropout from sklearn and Keras library itself to perform LSTM model. Load dataset of Finance Data of Goggle pay, Paytm, and Phone pe named as respectively.

STEP 2: Pre-process data set, in this step the data set is checked, and null values are replaced by Nan Values and dataset is normalized also the data is preprocessed with look_back =3 based on Volume by customers.

In this step the unique values from UPI banks are identified that includes Phone Pe, Google Pay, Paytm.

STEP 3: In this step we normalized transaction volume using MinMaxScaler to scale between 0 to 1. LSTM Function is defined to create and train data according to LSTM model, in this step to perform LSTM model the module is split into 80%-20% to train the model.

STEP 4: Define and compile LSTM model. The architecture consists of an Input LSTM layer with 50 units, Dropout Layer for regularization. The other LSTM layer with 50 units, second dropout layer. A dense layer with one unit for output.

```
Making predictions for Google Pay...
243/243 ━━━━━━━━━━━━━━━━━━━━ 2s 5ms/step
Predictions completed!
Making predictions for Paytm...
243/243 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step
Predictions completed!
Making predictions for PhonePe...
243/243 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step
Predictions completed!
```

Fig16: LSTM model predictions

STEP 5: In this step Train the LSTM model separately for each UPI app's data: Google Pay, Paytm, Phone pe.
STEP 6: Define a function for predictions and inverse transform the results to get actual values of data.
STEP 7: To forecast future transaction volumes for coming periods we have used the last available data point as the seed for generating future predictions.
Sum the forecasted transaction volumes for each app over the entire future period.

```
Total forecasted usage for Google Pay : 211916.58 Mn
Total forecasted usage for Paytm : 61100.44 Mn
Total forecasted usage for PhonePe : 174897.56 Mn
```

Fig17: Forecasted transaction volume for UPI banks

STEP 8: In this step we have determined the app that will be most used in future by applying LSTM model on finance data.

```
The most used app in future will be : Google Pay
```

STEP 9: Now let's visualize the future trends for determining the evolution of cashless payment applications based on Past actual Data. It determines the volume of customers using cashless payment application.
The visualizing is done using plot using python library matplotlib.
The visualizations will help us to get the evolution of cashless payment applications in future.

2]ARIMA model:
The methodology for forecasting future UPI transaction volumes for Google Pay, Paytm, and Phone Pe using ARIMA models involves the following steps:
STEP 1: Data Collection: The Finance Data is used that is extracted from government payment website that contains a UPI transactions volume of Phone Pe, Google Pay, Paytm Banks.
Let us look at the key columns in data we used to predict the growth of Cashless payment via UPI apps.

- Date: Transaction date in the format 'dd-mm-yyyy'.
- UPI Banks: Name of the UPI app.
- Volume (Mn) By Customers: Transaction volume in millions.

STEP 2: Data Pre-processing: Load the Dataset using Pandas Library. Once data is loaded convert a date column in datetime format. To follow proper time-series order sort data by date. Filter data for checking missing values in data.

```
Data Sample:
         Date                UPI Banks  Volume (Mn) By Customers  \
0  2017-04-01               Google Pay                  1.678545
30 2017-04-01                  PhonePe                  2.011986
60 2017-04-01  Paytm Payments Bank App                  0.513909
61 2017-04-02  Paytm Payments Bank App                  0.505561
1  2017-04-02               Google Pay                  1.423209

    Value (Cr) by Customers  Month  Year
0               3065.692692  April  2017
30              3812.536761  April  2017
60               527.198563  April  2017
61               515.399613  April  2017
1               3110.892611  April  2017
```

Fig18: Data Preprocessing step for NPCI dataset

STEP 3: ARIMA model specification: we defined ARIMA model with 3 parameters i.e. p, d, q was.

- p (autoregressive part): Number of lag observations included in the model.
- d (integrated part): Number of times the raw observations are differenced.
- q (moving average part): Size of the moving average window.

STEP 4: Fit the ARIMA model to the UPI transaction volume data for each application.
STEP 5: Let's forecast transaction volumes of each UPI application for future Time period.

```
Total forecasted usage for Google Pay : 241104.31 Mn
Total forecasted usage for Paytm : 52848.34 Mn
Total forecasted usage for PhonePe : 292326.08 Mn
```

Fig19: Volume of transactions per app
According to transaction volume forecasted above using ARIMA model it shows that the Phone pe has the highest volume of transactions than Goggle Pay and PAYTM according to ARIMA model.
STEP 6: Let us determine which app is predicted to be used more in future using ARIMA model:

```
he most used app in future : PhonePe
```

STEP 7: In this step we plotted the actual and forecasted transaction volumes for each UPI app to visualize the growth of application in future coming years.

3]SARIMA model:
This methodology determines the process of forecasting future UPI transaction volumes for Google Pay, Paytm, and Phone Pe by using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The following steps detail data preprocessing, model specification, forecasting, and visualization.

STEP 1: Data Collection: The Finance Data is used that is extracted from government payment website that contains a UPI transactions volume of Phone Pe, Google Pay, Paytm Banks.
Let us look at the key columns in data we used to predict the growth of Cashless payment via UPI apps.

- Date: Transaction date in the format 'dd-mm-yyyy'.
- UPI Banks: Name of the UPI app.
- Volume (Mn) By Customers: Transaction volume in millions.

STEP 2: Data Pre-processing: Load the Dataset using Pandas Library. Once data is loaded convert a date column in datetime format. To follow proper time-series order sort data by date. Filter data for checking missing values in data.
STEP 3: SARIMA model specification:
Specify the model SARIMA model parameters (p, d, q) for non-seasonal parameters (P, D, Q, s) for the seasonal part, where:

- p (autoregressive part): Number of lag observations in the model.
- d (differencing part): Number of times the raw observations are differenced.
- q (moving average part): Size of the moving average window.
- P, D, Q, s (seasonal components): Differencing, moving average terms, and the periodicity.
- For this study, the SARIMA parameters are set to (1, 1, 1) for the non-seasonal part and (1, 1, 1, 12) for the seasonal part.

STEP 4: Model Fitting: fit the SARIMA model to transaction volume of data for each UPI app. Sum the forecasted transaction volumes for each UPI app over the future period.
STEP 5: Let's forecast transaction volumes of each UPI application for future Time period.

```
Total forecasted usage for Google Pay : 244939.34 Mn
Total forecasted usage for Paytm : 53539.90 Mn
Total forecasted usage for PhonePe : 297386.68 Mn
```

Fig 20: transaction volumes of each UPI application

STEP 6: Let us determine which app is predicted to be used more in future using SARIMA model:

```
The most used app from will be: PhonePe
```

Fig21: predicted app to be used more in future using SARIMA model

STEP 7: In this step we plotted the actual and forecasted transaction volumes for each UPI app to visualize the growth of application in future coming year's by using SARIMA model.

I. ***Design Specifications:***
This study involves 2 datasets
Dataset1: App Reviews Data
Dataset2: Finance Data (NPCI Website)

a) Sentiment analysis:
Initially, let us see the design specifications for sentiment analysis. The design specification involves analysing users' sentiments from reviews of UPI applications (apps) i.e. Google Pay, Phone Pay and Paytm. By corelating these sentiments with finance data. The implementation step includes data preprocessing sentiment analysis and visualization.
>For data manipulation and analysis, we used pandas' libraries.
>We performed sentiment analysis using Vader sentiment analyser. It is used to find insights from a sentiment that are expressed in social media. We also used NLTK library for sentiment analysis on text data.
>For Visualization we used python Matplotlib library. To visual app reviews we used word cloud.
Loaded dataset of 3 UPI Apps, dropped unnecessary columns i.e. is `edited`, `developer response.
Using pie charts visualised distribution of sentiments classification also generated a word cloud to visualise most frequent terms in the reviews for each period and app.

6)Software Requirements:

- Python 3.6 +
- A setup of Jupyter notebook or Google Collab to code in Python is necessary.
- Pandas
- NLTK
- Matplotlib
- word cloud
- VADER lexicon from NLTK

7) Hardware requirements: Window/MacOS.

b) Machine learning/Deep Learning model:

Let us see the design specifications for implementing a machine learning and deep learning models i.e. using LSTM, ARIMA and SARIMA model for predicting growth of cashless payment application and prediction of App that will be more used in future using Finance Dataset**.**

1) Techniques: LSTM recurrent neural network model that can learn a long-term dependency and particularly suited for time series forecasting due to its ability to remember past information for longer periods. LSTM layers consist of LSTM units to process data sequentially.
2) Requirements: A UPI transaction volume data is collected from national payment website of India (NPCI). In data preprocessing we converted date column to datetime format. Also, we handle missing values from the data. For normalising the data, we used min Max scaler library and created a sequence with specified look back.
3) LSTM model Implementation: Defined and compiled the LSTM model, we split the data into train and test whereas the model is trained using training data and forecasted future values by using the trained model.
4) ARIMA model Implementation: This model is used for statistical analysis and to understand the time series data, The model uses a dependency between an observations and residual error from a moving average to lagged observations.
5) To fit ARIMA model by using an identified parameters that are p, d, q.
6) Forecasted future values to determine the growth of cashless payment applications. Along with prediction of most used application in future by using ARIMA model.
7) SARIMA model Implementation: This model extends ARIMA model which supports in univariate time series data within a seasonal component. This model accounts for its seasonality by differencing at a lag equal to the seasonal period. Seasonal autoregressive and moving average are added to the model.
8) We fit the SARIMA model to a finance data using identified parameters (p, d, q, P, D, Q, S). Using SARIMA model, we forecasted the future values to predict the evolution of cashless payment application in future along with predicting the most used Cashless payment application in future.
9) Libraries: we used pandas' library for data manipulation, NumPy for numerical operations, tensor flow and Keras for building and training the LSTM model, ARIMA model and SARIMA model. and matplotlib library for visualization purpose.
10) Environment and tools: Programming Language-Python. Jupiter notebook or any Python IDE is used.
11) Software/hardware requirements: Modern multicore processor. Sufficient storage for data sets and results full also a cross-platform operating system is used.

V. EVALUATION

This section represents a comprehensive analysis of the findings and predictions of study. By utilizing different techniques to assess performance and implications of sentiment analysis, LSTM, ARIMA, SARIMA models.

A) Sentiment Analysis Evaluation: The Sentiment Analysis is performed on the app reviews data, we analysed sentiments for PRE-COVID, POST-COVID and DURING COVID for Google pay, Phone Pe, and Paytm.
   Let us understand the public reviews with the help of word cloud for Phone Pe, Google Pay, Paytm.
   Google Pay-

Phone pe-

PAYTM-

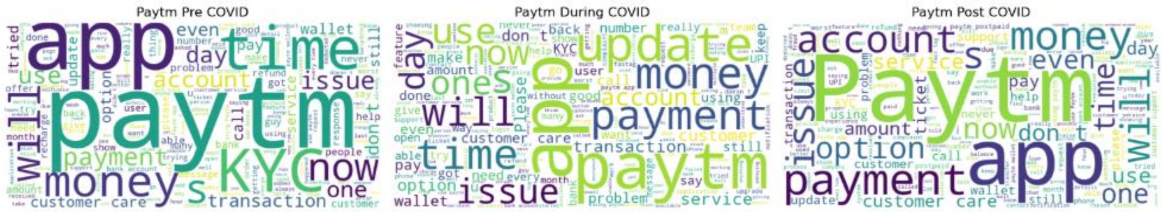


Fig22: Word Cloud of UPI app Reviews.

Now let us see the results achieved by analysing these sentiments by calculating a sentiment scores.

```
Google Pay Sentiments:
{'Google Pay': {'PRE COVID': {'Positive': 348, 'Negative': 202, 'Neutral': 21}, 'DURING COVID': {'Positive': 492, 'Negative': 3
53, 'Neutral': 41}, 'POST COVID': {'Positive': 579, 'Negative': 403, 'Neutral': 81}}}

PhonePe Sentiments:
{'PhonePe': {'PRE COVID': {'Positive': 372, 'Negative': 233, 'Neutral': 43}, 'DURING COVID': {'Positive': 197, 'Negative': 128,
'Neutral': 33}, 'POST COVID': {'Positive': 372, 'Negative': 211, 'Neutral': 51}}}

Paytm Sentiments:
{'Paytm': {'PRE COVID': {'Negative': 501, 'Positive': 465, 'Neutral': 47}, 'DURING COVID': {'Negative': 245, 'Positive': 226,
'Neutral': 16}, 'POST COVID': {'Negative': 245, 'Positive': 233, 'Neutral': 20}}}
```

Fig 23: sentiments analysis by calculating a sentiment scores

below are the results achieved after performing sentiment analysis by finding out sentiment score for every app. Let us first see for Google pay by visualising it through pie -chart:

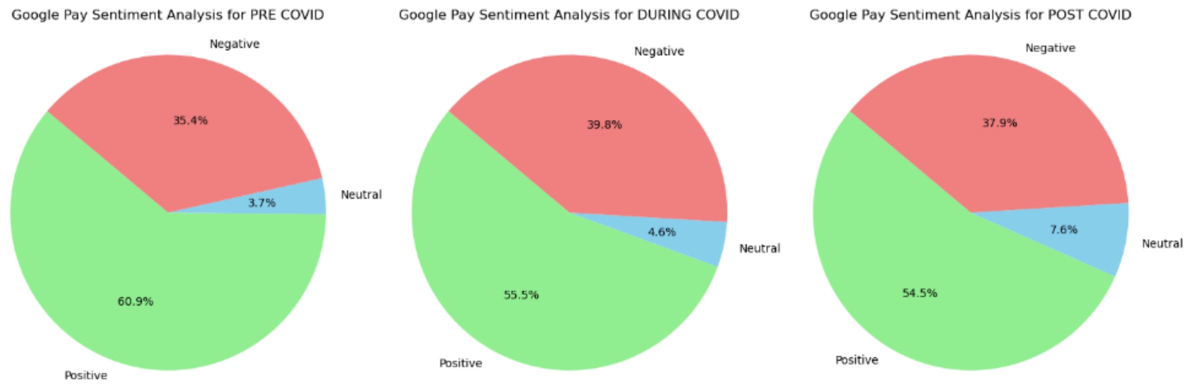


Fig24: Google pay Sentiments Visualisation.

In the above visualisation we can see that Google pay was used by public in all three periods before covid, during covid and post covid google pay is used and liked by people throughout since its inception.

Phone pe –

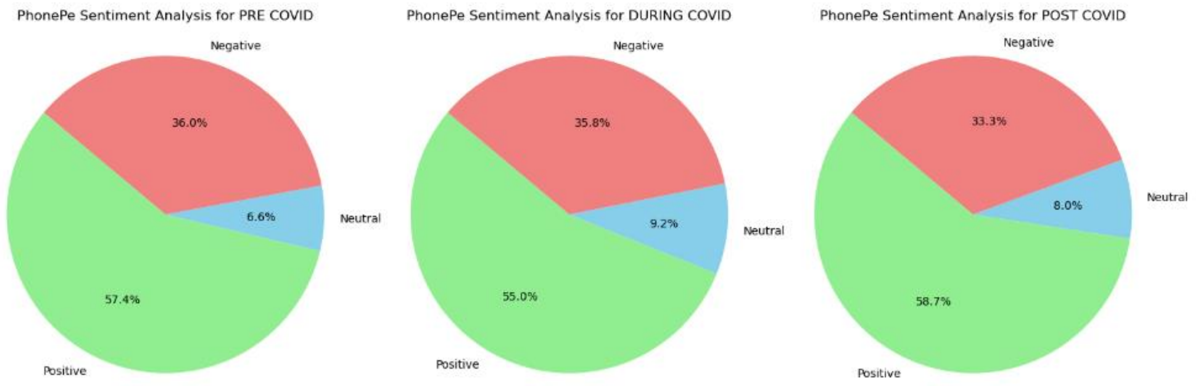


Fig25: Phone pe Sentiments Visualisation.

In the above visual we can observe that the Phone pe were used by people in all 3 periods and public reviews says that they contribute giving out excellent results along with google pay.

PAYTM-

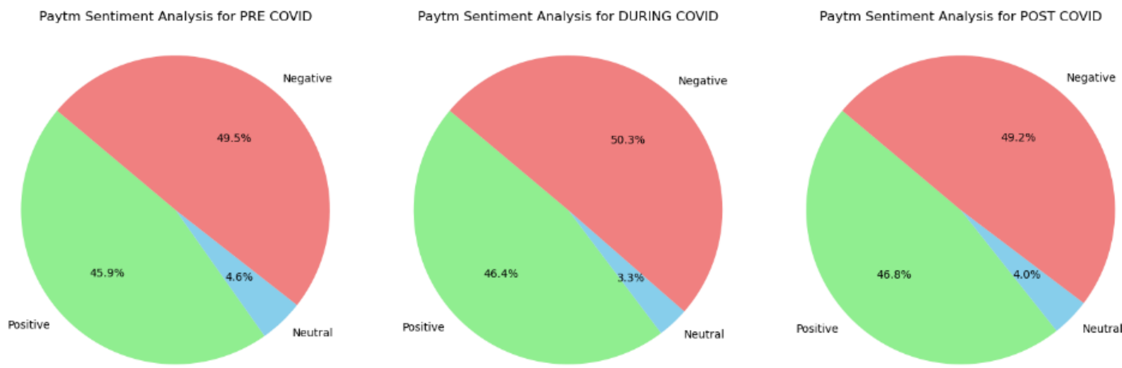


Fig26:Paytm Sentiments Visualisation.

As we can see in the above picture the PAYTM gives almost equal contribution based on public reviews as the positive and negative sentiments are almost equal for PAYTM during all 3 periods.

The sentiment analysis reveals a distinct sentiment distribution for Google Pay phone pay and Paytm for different periods recovered during covid and post covid, Pie charts and word clouds helps us to get a visual insight from the sentiment trends showing the prevalence of positive neutral and negative sentiments.

Implications: The methodology of sentiment analysis provides the applicability of VADER lexicon for the app review data. It provides a framework similar research exploring user sentiment trends and their implications. Practitioners can leverage this methodology of sentiment analysis to understand the user feedback better enhance user experiences and address the negative sentiments proactively.

B) LSTM Model Evaluation:

The LSTM model trained for Google Pay, Phone Pe, Paytm finance data gives us a good fit to actual Historical data. The models were used to observe and predict the growth of cashless payment application from 2025 onwards. The predicted values indicate the growth trend for each UPI app.

Using LSTM model, we predicted the cashless payment app that will be more used by people in future.

```
The most used app in future will be : Google Pay
```

Fig27.Prediction of most used app in future using LSTM model

Let us see the mean calculation forecasted for each app using LSTM.

```
Total forecasted usage for Google Pay : 211916.58 Mn
Total forecasted usage for Paytm : 61100.44 Mn
Total forecasted usage for PhonePe : 174897.56 Mn
```

Fig 28. Mean calculation forecasted

By observing above calculation, we predicted the trends to observe growth of every cashless payment application based on historical data by using LSTM model.

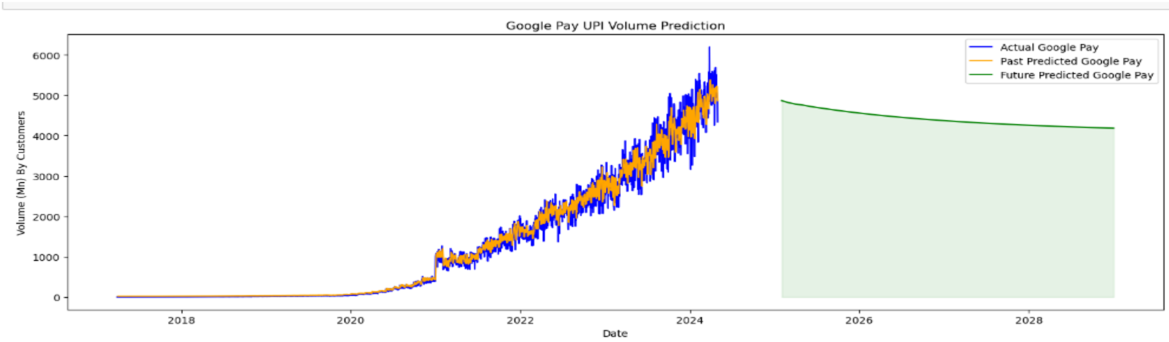


Fig29. LSTM model predictions of the actual, past predictions and Future Predictions for google pay

The above graph tells us that using LSTM model predictions of the actual and past predictions for google pay are predicted accurately and future prediction tells us that google pay has leveraging growth in cashless payment app.

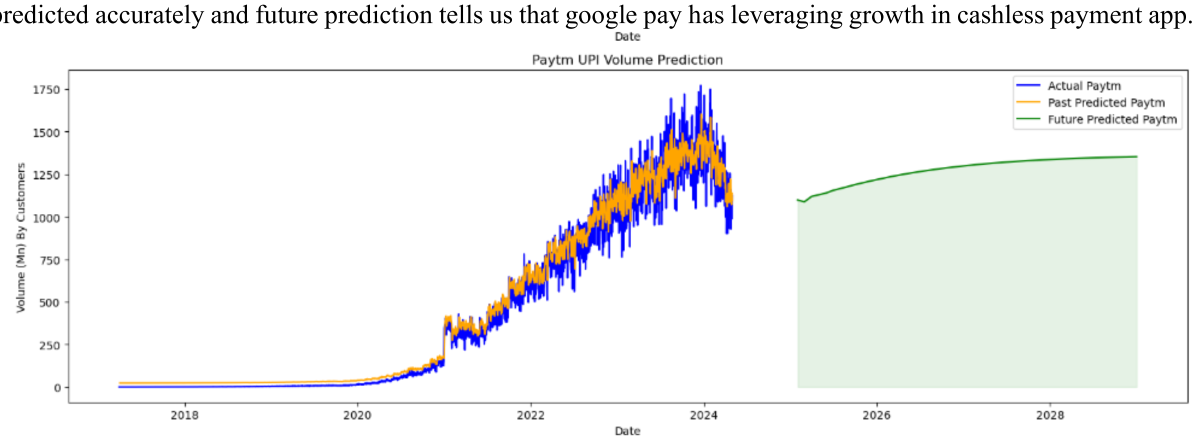


Fig30. LSTM model predictions of the actual, past predictions and Future Predictions for Paytm pay

PAYTM proves to be also evolving cashless payment app in future market by using LSTM model.

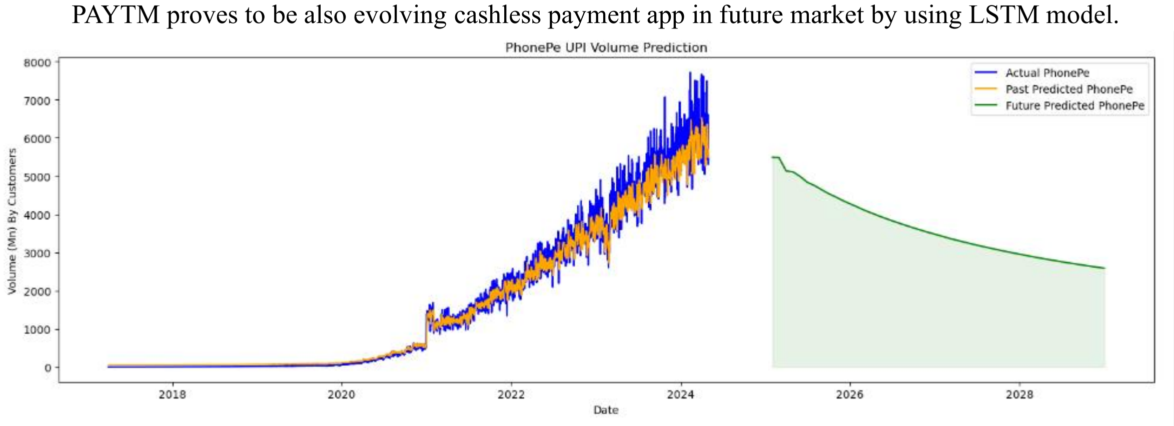


Fig31. LSTM model predictions of the actual, past predictions and Future Predictions for Phone pe

By using LSTM model, the predictions for phone pe indicates that the transaction volume of phone pe is not evolving in future, whereas the Phone pe should take care how they can leverage their market value and keep growing in market after these predictions from LSTM model.

The LSTM is used for time series forecasting in finance data that demonstrates its efficiency in capturing temporal dependencies and making accurate predictions.

C) ARIMAMODEL:

Using Arima model we tried to forecast the growth of cashless payment apps and predict the app that will be most used in future.

Let us see forecasted transaction volume of all 3 applications.

```
Total forecasted usage for Google Pay : 241104.31 Mn
Total forecasted usage for Paytm : 52848.34 Mn
Total forecasted usage for PhonePe : 292326.08 Mn
```

The most used application in future according to ARIMA model prediction is:

```
The most used app in future : PhonePe
```

Let us see trends in future forecasted values

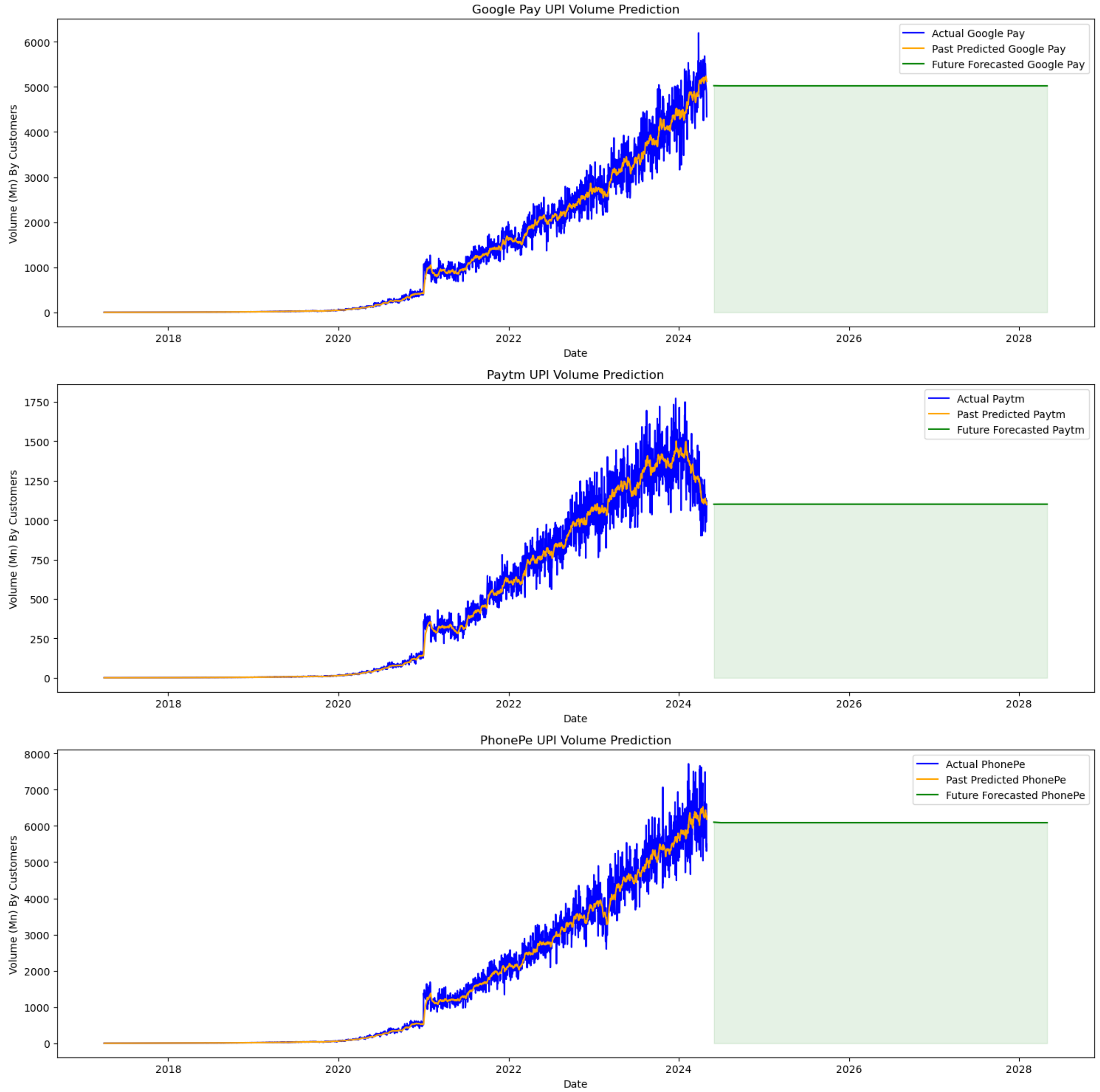


Fig32. ARIMA model predictions of the actual, past predictions and Future Predictions for Google Pay, Paytm and Phone pe

According to ARIMA model prediction the Google pay shows us the stable growth of cashless application in future. The PAYTM will be growing as a normal app in market compared to other 2 applications in future according to its trend in plot above.

As we can see using ARIMA model the Phone pe gives us highest volume of transactions forecasted for future and phone pe proves to be a most used app in future according to official government data.

D) SARIMA model:

Let us the see evolution of cashless payment apps using SARIMA model by capturing the seasonal trends in data which helps us to provide more accurate results.

Let us observe the forecasted transaction volume of apps using SARIMA model

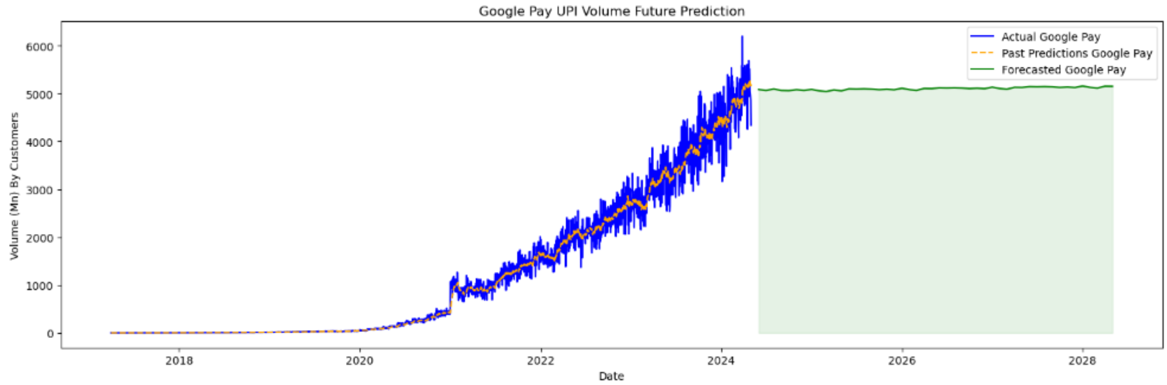
```
Total forecasted usage for Google Pay : 244939.34 Mn
Total forecasted usage for Paytm : 53539.90 Mn
Total forecasted usage for PhonePe : 297386.68 Mn
```

Let us see the most used app predicted by using SARIMA model:

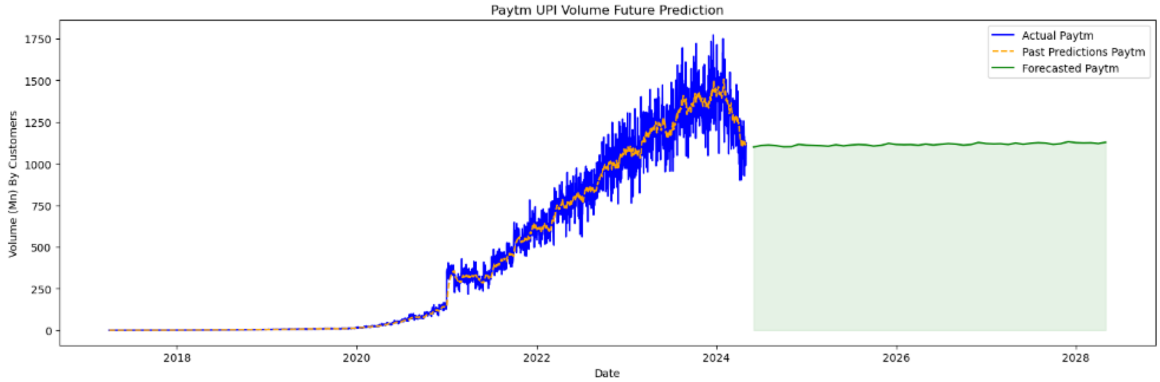
```
The most used app from will be: PhonePe
```

The SARIMA model and ARIMA model gives us the most used app as Phone Pe

Let us observe the future evolution of apps using SARIMA model:



Google Pay give us the stable growth in use of application for future according to SARIMA model.



The Paytm UPI bank shows the normal usage of app by the people in future.

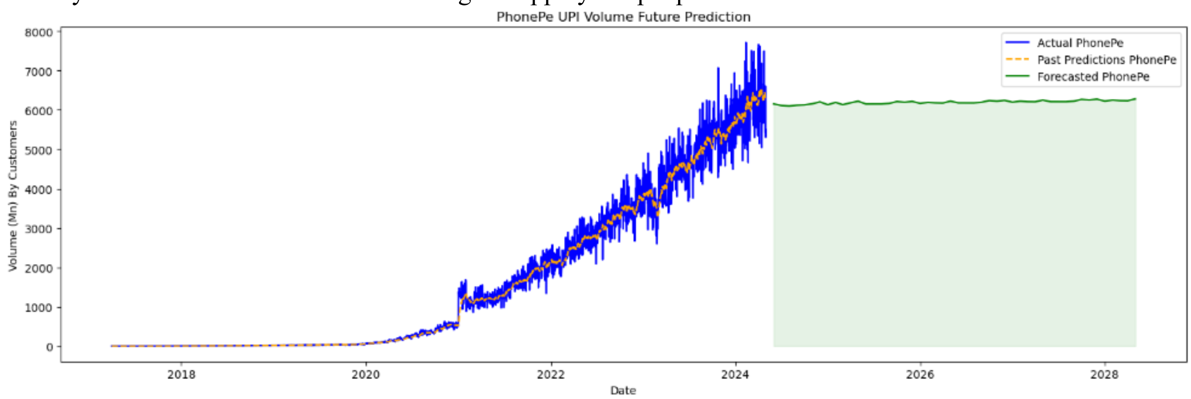


Fig33. SARIMA model predictions of the actual, past predictions and Future Predictions for Google Pay, Paytm and Phone pe

Phone pe shows us the highest growth of application users in future compared to other two.

According to ARIMA and SARIMA model the evolution of cashless payment application is predicted, and PHONE PE proves to be the most used app in future. Using LSTM model, the Google pay application has the better growth in future. The cashless payment proves to be evolving via apps. We can see the sentiments in during covid, pre covid and post covid and after affecting the covid the cashless payment apps kept evolving mostly the PHONE PE and Google Pay. We can see the actual growth of cashless payment application evolving in India by analysing sentiments and examining machine learning and deep learning models on UPI data from National payment UPI website.

VI. Conclusion And Discussion.

1)**Sentiment analysis:**

Pre-Covid: After analysing sentiments for Google Pay, Phone Pe and Paytm there was positive response for all three apps in PRE COVID period.

During -Covid: There was positive response from people especially for Google Pay and Phone pe whereas Paytm received the number of significant negative sentiments.
Post -Covid: According to sentiments, Google Pay, Phone Pe continued to be a remarkable UPI app which evaluated the growth of cashless payment in India instead Paytm is still not much used by the common public.
2)**Transaction Volume Of UPI:** Evolution based on UPI transaction volume data, UPI apps from 2017 to 2024 the Google Pay and Phone pe exhibited a constant growth in transaction volumes while Paytm show a slow growth in future.
3) **Future predictions for evolution of cashless payment apps in India:**
LSTM model: Using LSTM model the Google Pay and Phone Pe shows continuous growth of cashless payment app in market.
ARIMA model: Forecasted the phone pe has the highest transaction volume for cashless payments, compare to Google Pay and Paytm.
SARIMA model: Predicted the growth of cashless payment via phone pe, Google Pay and Paytm. For coming year's using SARIMA model the phone pe and google pay is showing us the high evolution of cashless payment.
4)Insights Gained:
After sentiment analysis the positive sentiments and high transaction volumes for Google Pay and phone pe highlights the successful market for cashless payment evolution in India.
Impact of COVID-19 on cashless payment apps heightened engagement and reliance on these apps according to sentiment analysis
5)Future Growth: The forecasted models indicate a promising future for digital payments in India with continuous evolution in future and integration in everyday transactions.
6) Future work And Discussion:
A future study can compare the growth and user sentiments of UPI based apps with other digital payment methods.
The evolution of cashless payment in India could be implemented by using both the finance data from npci and by analysing a sentiment from app.
The adoption and growth of cashless payment systems can be studied according to changes in government policies and regulations.
While using a hyper tuning parameter in machine learning models for future prediction of cashless payment app would provide more accurate results. The model accuracy can be improved by using Hyper parameter tuning.
Current situation from national news tells us that the evolution of Paytm is dropped whereas, our model proves to predict the actual situation for Paytm app.
7)Relevance to stakeholder: This research helps to get actionable Insights into a user preferences and growth trends with helping companies to plan a finance according to market. Financial institutions can use this data to develop partnerships and integrate a digital payment solution. The insights from this research can guide policymakers ensuring security in digital payments ecosystem.
8) Data driven decision making: By taking an advantage of transaction volume forecast to understand market trends and customer needs, the finance institution can take informed decision on product development and market expansion

In conclusion, the research presents that digital payment applications have an important impact on the finance industry. By analysing public sentiments and growth trends using UPI finance data it provides valuable insights for stakeholders to take informed decisions according to forecasted cashless payment ecosystem.

VII. REFERENCES

1. Ganapathyraman, S., Suresh, S. and Thomas, T. (2023) *A study on users' opinion towards unified payment interface (UPI) transactions*, *SSRN*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4566690

2. A., M. and S., G. (no date) *India's Digital Payment Landscape – An Analysis*, *International Journal of Case Studies in Business, IT and Education (IJCSBE)*. Available at: https://www.supublication.com/index.php/ijcsbe/article/view/847

3. Balakrishnan, M. (2023) *The unified payment interface and the growth of digital payments in India: An analysis*, *Latest TOC RSS*. Available at: https://www.ingentaconnect.com/content/hsp/jpss/2023/00000017/00000003/art00004

4. *[PDF] a study of academic performance using Random Forest , artificial neural network , naïve Bayesian and logistic regression | semantic scholar*. Available at: https://www.semanticscholar.org/paper/A-Study-of-

Academic-Performance-using-Random-Forest-Ulinnuha-Sa%E2%80%99dyah/509053ce692561ae9fb402412c7d1d8b94f6547e.

5. Person, Piyush, Z. and Gupta, R. (2024) *Digital Revolution and innovation in the banking and finance sectors |, Taylor & Francis*. Available at: https://www.taylorfrancis.com/chapters/edit/10.1201/9781032618845-4/digital-revolution-innovation-banking-finance-sectors-piyush-gupta-zehra-raza.

6. Glez-Peña, D. *et al.* (2013) *Web scraping technologies in an API world*, *OUP Academic*. Available at: https://academic.oup.com/bib/article/15/5/788/2422275.
7. Venkatakrishnan, S., Kaushik, A. and Verma, J.K. (1970) *Sentiment analysis on Google Play Store data using Deep Learning*, *SpringerLink*. Available at: https://link.springer.com/chapter/10.1007/978-981-15-3357-0_2.
8. *How do users like this feature? A fine grained sentiment analysis of App Reviews | IEEE conference publication | IEEE xplore*. Available at: https://ieeexplore.ieee.org/document/6912257/.
9. Joy, A., Rejikumar, G. and Dhanya, M. (2019) *Payment app's revamping, Perspectives from text mining analysis*, *EUDL*. Available at: https://eudl.eu/doi/10.4108/eai.18-12-2018.2285156.
10. Mahesh A. 1, 2 and Ganesh Bhat S. (no date) *View of a systematic review and research agenda of digital payment system with reference to unified payment interface*. Available at: https://www.supublication.com/index.php/ijmts/article/view/582/444.
11. *(PDF) 'ASM's International E-Journal on ongoing research in management and it' E- role of private channel members as alternative means of payment within the common framework of Retail Payments System Regulation*. Available at: https://www.researchgate.net/publication/275154061_ASM's_International_E-Journal_on_Ongoing_Research_in_Management_and_IT_E-_Role_Of_Private_Channel_Members_As_Alternative_Means_Of_Payment_Within_The_Common_Framework_Of_Retail_Payments_System_Regulati.
12. Khatwani, R., Mishra, M. and Bedarkar, M. (2023) *Growth and penetration of digital payments in India: Regional analysis*, *AIP Publishing*. Available at: https://pubs.aip.org/aip/acp/article-abstract/2954/1/020013/2925115/Growth-and-penetration-of-digital-payments-in.
13. Goel, N. and Nath, V. (no date) *An exploratory study on digital payment systems and its impact* ... Available at: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3562948_code3635775.pdf?abstractid=3562948&mirid=1.