Paytm reviews sentiment analysis and topic modelling with recommendation model in machine learning

Group 2- Amit Kumar Jha, Sweta Tiwari

Background

The rise of web 2.0 is reshaping the social media landscape. Not only are people using online social media to interact, share content, and express their personal opinions with others, but businesses can also utilise social media for communication, analysis, and enhancement of their products and services.Every day, the number of people using social media grows, and it is anticipated that by 2019, there will be 2.77 billion users worldwide.Social media is a goldmine of raw and unrefined data, and advances in technology, particularly in artificial intelligence and machine learning, have enabled the data to be processed and converted into meaningful information that may help almost any business.Sentiment analysis (also known as opinion mining) is a natural language processing (NLP) technique for determining the positivity, negativity, or neutrality of data. Sentiment analysis is frequently used on textual data to assist organisations in tracking brand and product sentiment in consumer feedback and better understanding customer demands. The Recommendation system helps businesses to find potential customers whether a particular product should be recommended or not. The recommendation system helps customers to sort their choice of product based on the reviews and rating available on the platform.

Motivation

Sentiment classification and topic modelling is becoming a crucial tool for monitoring and understanding user sentiment as they share their opinions and feelings more openly than ever before.Businesses can understand what makes customers happy or frustrated by automatically evaluating customer feedback, such as comments in survey replies and social media dialogues. This allows them to customise products and services to match their customers' demands.

This project examines analysis of reviews posted on google play store in order to provide a better understanding to the respective organisation. Sentiment analysis is a method of extracting, converting, and interpreting opinions from a text and categorising them as positive, negative, or natural sentiment using Natural

Language Processing (NLP). Topic modelling is an unsupervised machine learning algorithm that can generate a number of clusters of topics discussed in frequent numbers in collection of texts.In the business world, sentiment research can be a game-changer for total brand rejuvenation. The capacity to exploit unstructured data for actionable insights is critical to running a successful business utilising sentiment data. Sentiment analysis can be useful for businesses in many ways like from building business intelligence systems to get competitive advantage and enhancing the customer experience.

ML innovation:

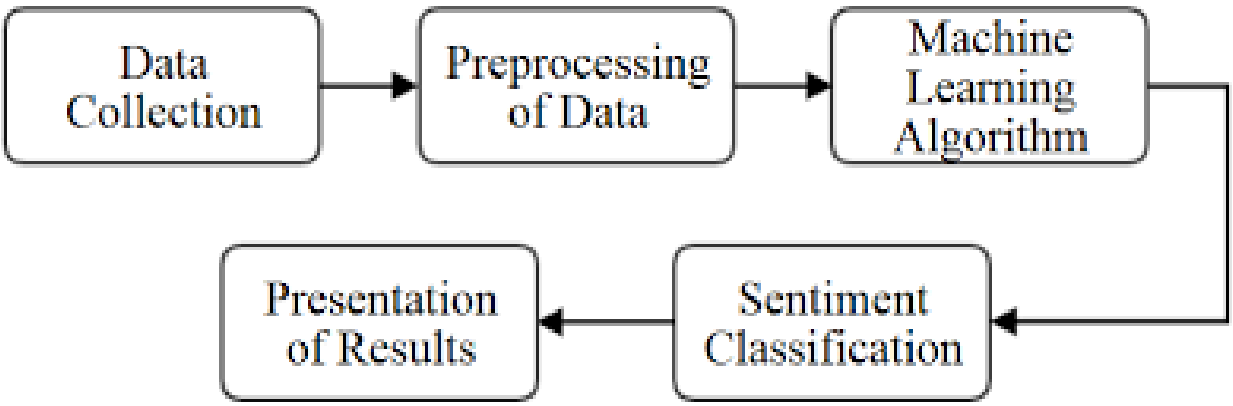
Building reviews based recommendation system: The review-based recommendation system was created with the primary goal of extracting relevant information from a user's textual review of a product. This is how machine learning and natural language processing can be combined. We have implemented followings to get sense of how textual reviews are useful for businesses likes paytm

- Natural Language processing
- Emotional analysis
- Topic modeling
- Classification model using Naive Bayes classifier

Sketch of system design :

For sentiment analysis:

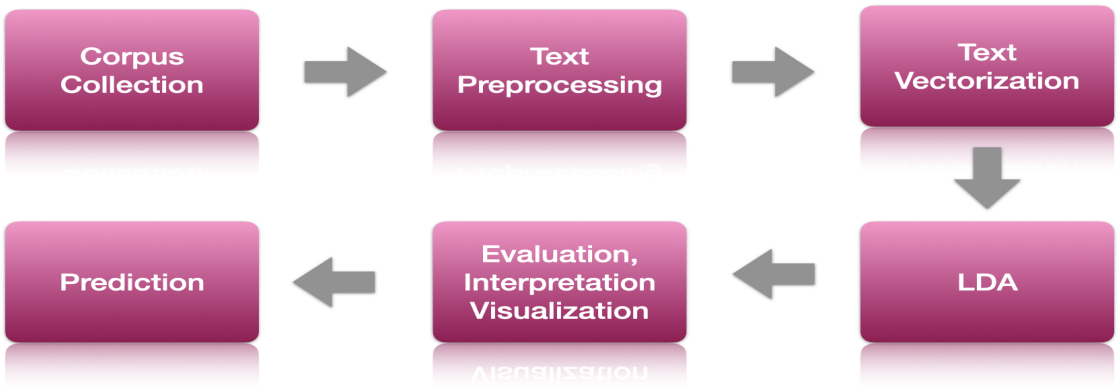
- ➢ Data collection: We have scraped around 1500 reviews posted related to FinTech giant paytm from google play store.
- ➢ Preprocessing of Data: Removal of stop words, stemming, lemmatization, punctuation removal, number removal and also symbol or emoji removal. We have used the NLTK library from the Python language.
- ➢ AI/ML algorithm: NLP
- ➢ Sentiment classification : Positive , negative and neutral
- ➢ Emotion analysis : anger, fear, happiness, sadness, disgust etc.
- ➢ Presentation: Visualization of frequency of sentiments, word cloud, topics



For Topic modelling :

- ➢ Corpus collection: Document creation using reviews texts
- ➢ Text preprocessing :Removal of stop words, stemming, lemmatization, punctuation removal, number removal and also symbol or emoji removal
- ➢ Vectorization : Creation of vectors of texts
- ➢ Implementation of **Latent dirichlet allocation(LDA)** algorithm to create topics of having high probability
- ➢ Visualization of topics , terms
- ➢ Prediction

Topic Modeling Pipeline



Process and results :

- First we imported the required libraries like pandas , matplotlib, sklearn, word cloud, request and nltk etc

```
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import numpy as np # linear algebra
import seaborn as sns # plotting
import matplotlib.pyplot as plt # plotting
%matplotlib inline
import os # acce
from wordcloud import WordCloud
import nltk
from nltk.corpus import stopwords
from nltk import sent_tokenize, word_tokenize
from wordcloud import WordCloud, STOPWORDS
from collections import Counter
from nltk.tokenize import RegexpTokenizer
from stop_words import get_stop_words
import re
import spacy
```

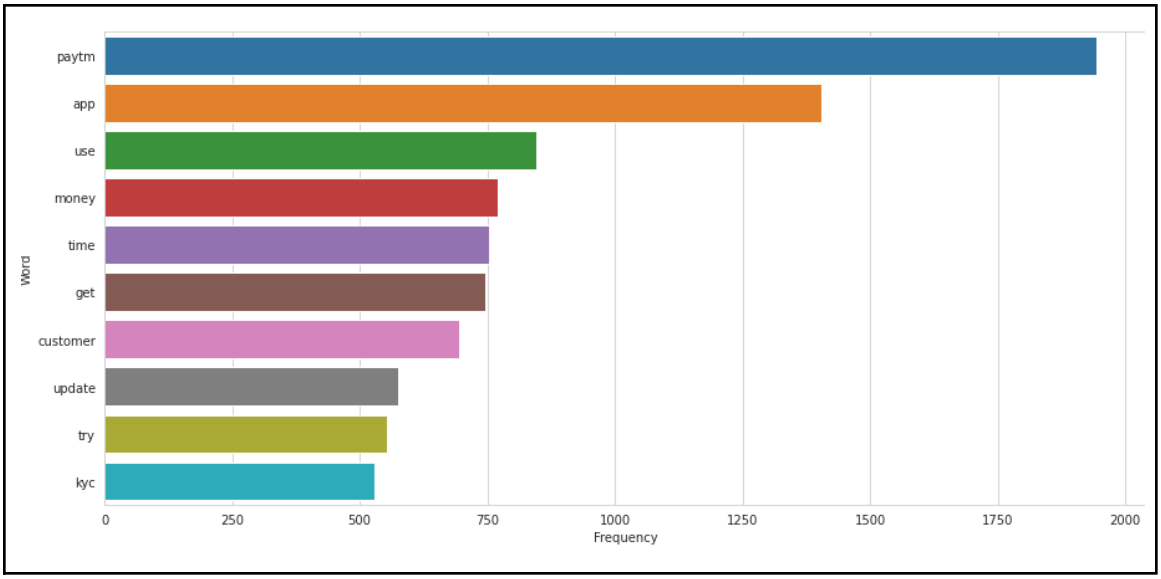
- Data import : using pandas , we have imported our data set of reviews of paytm which we have extracted from google play store
  Below figure shows our dataset , we have 1500 reviews in total and 5 variables

```
data
```

	Unnamed: 0	userName	rating	date	title	review
0	0	rohit wangde	5	7/17/2021	Very helpful	i m customer of paytm long back but got discon...
1	1	Shubham 'the Hun	4	1/15/2020	Frequent Updates	This hassle of updating the app every other da...
2	2	KkGohel	5	1/4/2020	Problem about upgrading the application	Why i should have upgrade the application if I...
3	3	Sailendra Kumar Sahoo	3	7/3/2020	Good App but bad customer experience	This is one the best app for A-Z transactions....
4	4	dnowncsnxh	1	5/11/2019	KYC!!	When i tried doing KYC last year when they sta...
...	...	...	...	...	...	...
1495	1495	NiteshThat	5	7/7/2018	Not working well anymore	It takes a long time to put trnafer money from...
1496	1496	jon_strickson	5	6/29/2018	Delay and stuck of money	Money is cut off from the bank acout but not p...
1497	1497	Rose4Firoz	5	4/8/2018	Paytm app touch not working	After the recent update the main screen contai...
1498	1498	Dk_43	1	5/29/2018	Pathetic service	These people might be spending millions luring...
1499	1499	this is fraud app	2	5/5/2018	Customer service so insensitive	I got debit card of Paytm but it is of no use ...

1500 rows × 6 columns

- After using standard text preprocessing like removal of stop words, removal of punctuation, removal of numbers and symbols we have created frequency of words discussed in corpus of reviews, here we can see as company is paytm customers used paytm in higher numbers and then some important like KYC, money are discussed.

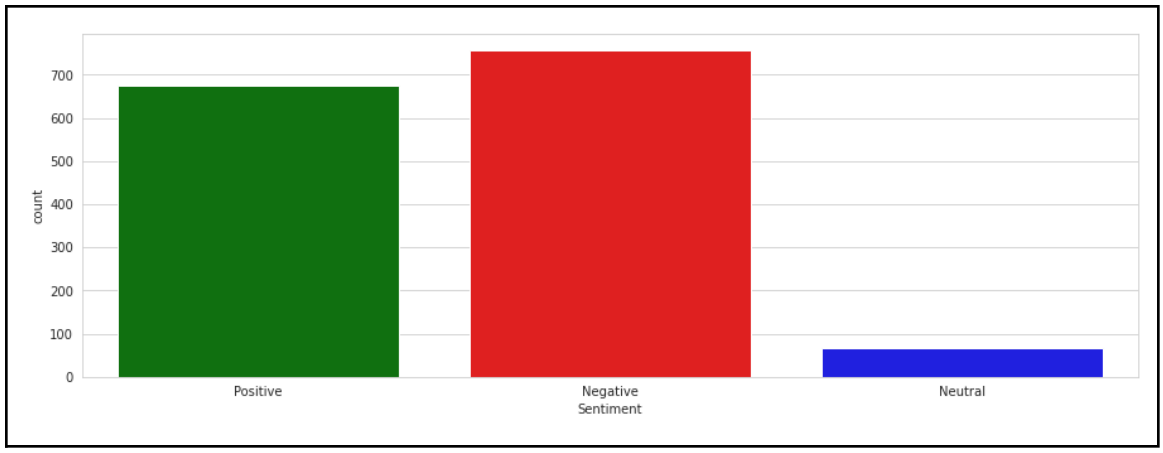


- After creating frequency of words , we created an interactive word cloud to see visualization of most frequent words in the reviews corpus.Word clouds are a low-cost alternative to coding for evaluating text from online surveys, and they're also significantly faster.

- Then we implemented sentiment analyzer to calculate polarity score of each review.Polarity is floating point which lies in the range of **[-1,1]** where close to 1 means positive statement and close to -1 means a negative statement

	review	Polarity
0	i m customer of paytm long back but got discon...	0.9774
1	This hassle of updating the app every other da...	-0.6137
2	Why i should have upgrade the application if I...	-0.7789
3	This is one the best app for A-Z transactions....	0.9858
4	When i tried doing KYC last year when they sta...	-0.9378
...	...	...

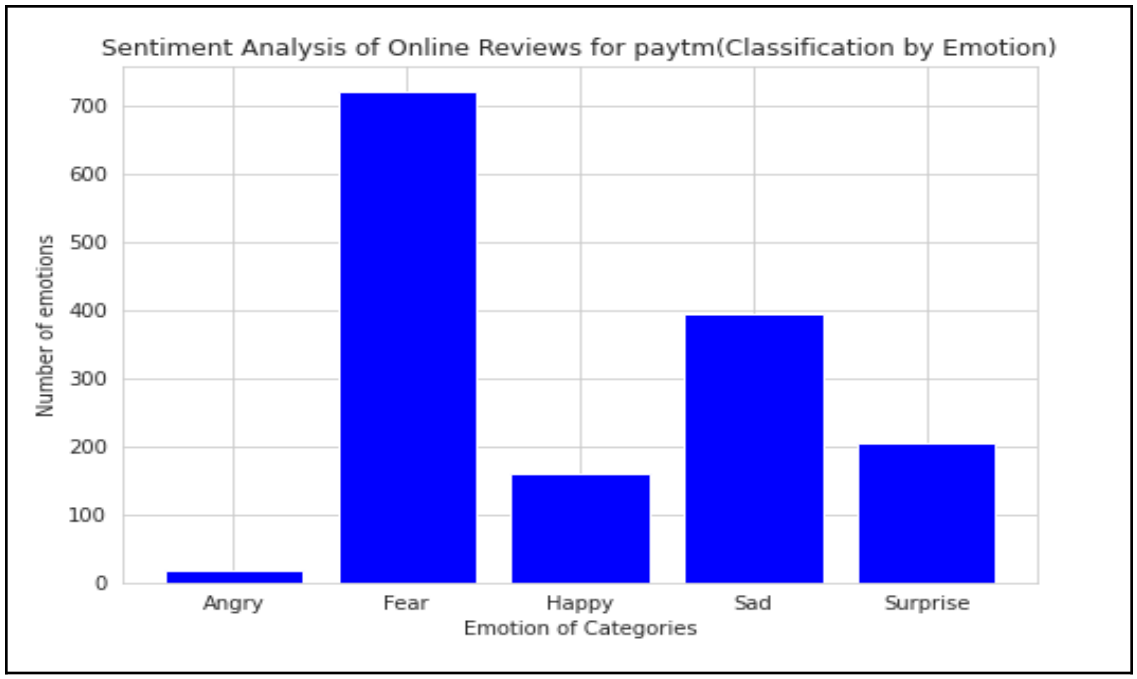
- Based on the polarity score, we have categorised the sentiments into three categories positive for score greater than 0.05 and negative for score less than -0.05 , else all sentiments come under neutral category.



- Customer experience can be analyzed if we can see the emotions they are showing over the social media platforms, so we also performed emotions analysis and successfully create frequency of various emotions like Fear, surprise , Happy, Sad etc . We used text2emotion library for this purpose.

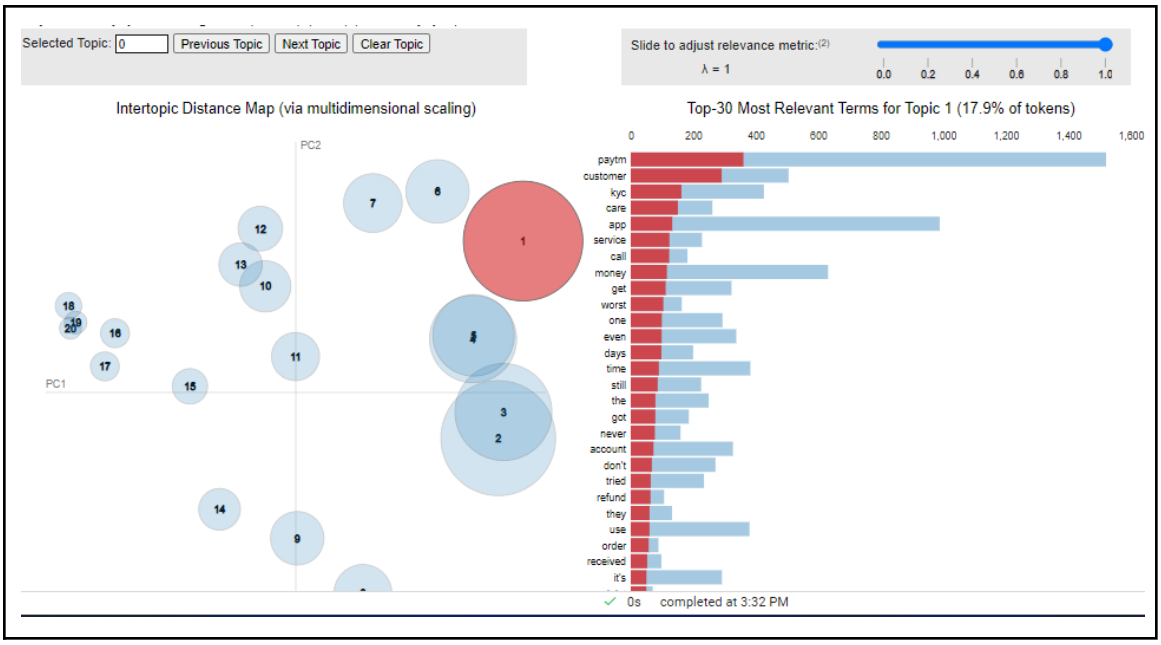
	review	Polarity	Sentiment	Emotions
0	i m customer of paytm long back but got discon...	0.9774	Positive	Fear
1	This hassle of updating the app every other da...	-0.6137	Negative	Fear
2	Why i should have upgrade the application if I...	-0.7789	Negative	Surprise
3	This is one the best app for A-Z transactions....	0.9858	Positive	Fear
4	When i tried doing KYC last year when they sta...	-0.9378	Negative	Fear
...	...	...	...	...

- We also created visualisation of emotions because pictures are always more elaborative than words.Here we can see that customers are in fear more than the happy and surprises.Emotion analytics can gather text data from a variety of sources in order to assess subjective data and comprehend the emotions underlying it.



- Below figure shows examples of topic with terms discussed in each topic , the number shows probability of each term , these topic were created with the help of LDA algorithm

- After that , we have visualised various topic clusters and calculated intertopic distance with the help of pyLDAvis package in python. When we click on particular cluster like in below in cluster 1 we can observe frequency of each term used in the particular cluster



Recommendation Model building using Naive Bayes Classifier

After successful analysis of sentiment and topics discussed in reviews now we have implemented a supervised machine learning algorithm to predict whether based on reviews paytm should be recommended or not recommended. First we gave the label to categorise based on polarity score so if the score is greater than 0.05 then we have labeled 1 means recommended otherwise 0 means not recommended. For classification and prediction we have used naive bayes classifier algorithm and got 70% accuracy with an F1 score of 68%.
Package used for ML is **sk learn**

	review	Label
0	i m customer of paytm long back but got discon...	1
1	This hassle of updating the app every other da...	0
2	Why i should have upgrade the application if I...	0
3	This is one the best app for A-Z transactions....	1
4	When i tried doing KYC last year when they sta...	0
...	...	...

- Now we tested the model on a testing dataset, took one negative review from a random index and tried to check if the model is predicting a good or significant result or not and below figure says yes model predicted perfectly fine. Based on that if someone posts a negative review based on that paytm shouldn't be recommended and for positive reviews model predicts paytm should be recommended.

```
# trying to make a prediction using this review
# this is an example for negative review
negative_example = data['review'][1]
negative_example
```

```
'This hassle of updating the app every other day or probably twice in every week is really tiring , it's convenient if you have WiFi but when you are in any publ
ic place or for that matter the majority of Internet users in this country depend on Mobile Data rather than WiFi and the increasing size of the application and
sometimes due to high usage of data in a particular area/location, network server tends to be slow and results in consumption of extended duration of time thus ,
hassle . I know some points made here are beyond your control such as network providers and stuff , but you being one of the major payment gateway in India and h
aving a renowned presence in market should design the app according to its users within this land mass , make it quicker and smoother . \nThe other day it was la
te night , my phone was gonna die and when I had to use paytm as I didn't have any cash then , it consumed hell lotta time,eventually my battery flushed out help
lessly I had to look for atm . \n\...'
```

```
# vectorize the positive_example
positive_example_vec = count_vectorizer.transform([positive_example])
# make prediction
nb.predict(positive_example_vec)[0]
```

```
'0'
```

- Predicting with some random reviews as an input and here is the output, the review clearly shows positive polarity and model correctly predicted app should be recommended.

```
example='This app is really awesome and I liked the most'
example
```

```
'This app is really awesome and I liked the most'
```

```
negative_example_vec = count_vectorizer.transform([example])
nb.predict(negative_example_vec)[0]
```

```
'1'
```

Business potential :

Many companies are attempting to launch their FinTech, there is a tremendous chance for established payment apps like PayTm to maintain market share. Sentiment analysis in business aids in assessing current and prospective customer's

perceptions of all of these elements. Keeping the bad emotions in mind, you can build more appealing branding approaches and marketing strategies to help your brand transition from mediocre to fantastic.To do so, customer experience & segmentation will play a critical role, since the best approach to maintain market share is to get to know your customers better! Customer impression on various social media platforms helps businesses to divide their customers on the basis of several factors. We can create an app that on the basis of reviews provided by customers on various social media platforms can build a recommendation system of various apps. We can build comparative analysis of various apps on the basis of ratings provided by customers. Nowadays due to the large number of fintech applications available, customers are confused which app is having good reviews so we are thinking of developing an app called finalytics using flask on which we'll integrate the recommender system to give the recommendations based on reviews.

Limitations:

We can implement different supervised learning models like XGBoost, Logistic regression to check accuracy and also we can increase the number of data points or reviews to make better predictions. Also we can elevate to multiple digital payment apps to see how customers are shifting from one app to another.