ANALYSE THE CUSTOMER REVIEW FOR NEWLY LAUNCHED MOBILE USING: THE TOPIC MODELLING APPROACH

Present By:
Asish Kumar Nayak -21MB0013

AGENDA

Introduction

Problem Statement

Project progress

- Data summary
- Data analysis
- Topic Modelling
- Text classification

INTRODUCTION

E-commerce has revolutionized the way we shop. That phone you've been saving up to buy for months? It's just a search and a few clicks away. Items are delivered within a matter of days (sometimes even the next day!).

For online retailers, there are no constraints related to inventory management or space management They can sell as many different products as they want. Brick and mortar stores can keep only a limited number of products due to the finite space they have available.

But online shopping comes with its own caveats. One of the biggest challenges is verifying the authenticity of a product. Is it as good as advertised on the e-commerce site? Will the product last more than a year? Are the reviews given by other customers really true or are they false advertising? These are important questions customers need to ask before splurging their money.

PROBLEM STATEMENT

- ❑ Consumers may learn a lot from online reviews of products.
- ❑ Online reviews may be used by vendors to get feedback from customers on the goods or services they are selling.
- ❑ Yet, because the volume and depth of these online evaluations are sometimes overwhelming.

Topic Modelling

- ❖ Let customers to rapidly go through the reviews' most important points without having to read all of them.
- ❖ Assist the merchants/retailers in obtaining consumer input in the form of topics

Project Objective

- ❑ In our case, we have thousands of online reviews for the newly launched mobile.
- ❑ Our aim here is to extract a certain number of groups of important words from the reviews.
- ❑ These groups of words are basically the topics which would help in ascertaining what the consumers are actually talking about in the reviews.

Programming language and algorithm

- ➢ Python Language
- ➢ Here we'll work on the problem statement defined above to extract useful topics from our online reviews dataset using the concept of Latent Dirichlet Allocation (LDA).

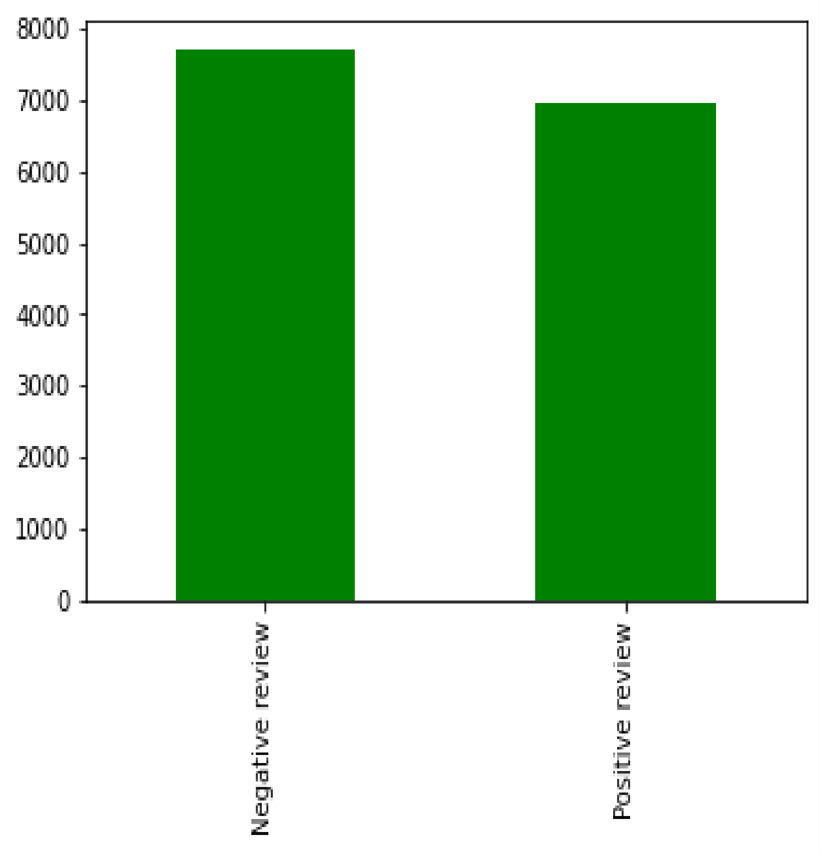
PROJECT FRAME WORK

DATA COLLECTION	DATA CLEANING	TF-IDF	FIND THE NUMBER OF CLUSTER	TOPIC MODELLING
Product review	Data preprocessing	Document frequency	Elbow method & gap statistic	LDA approach

DATA

Customer review about the mobile labeled with positive or negative sentiment.

Features	Count	Data type
Sentiment	14675	Integer
Reviews	14675	Object



Negative	7712
Positive	6393

DATA CLEANING

Python Code:

```
from cleantext import clean
clean(data['review'], no_emoji=True)
#------------------------------------------------------------------------------
# function to remove punctuation
import string
def clean_text(text ):
    delete_dict = {sp_character: '' for sp_character in string.punctuation}
    delete_dict[' '] = ' '
    table = str.maketrans(delete_dict)
    text1 = text.translate(table)
    textArr= text1.split()
    text2 = ' '.join([w for w in textArr if ( not w.isdigit() and  ( not w.isdigit() and len(w)>3))])
    return text2.lower()
data['review'] = data['review'].apply(clean_text)
#------------------------------------------------------------------------------
#Let us pre-process the data
from nltk.corpus import stopwords
stop_words = stopwords.words('english’)

# function to remove stopwords
def remove_stopwords(text):
    textArr = text.split(' ')
    rem_text = " ".join([i for i in textArr if i not in stop_words])
    return rem_text

df = data
# remove stopwords from the text
df['review']=df['review'].apply(remove_stopwords)
```

Remove the emoji from the reviews

Remove the Punction from the reviews

Remove the Stop words (English) from the reviews

WORD CLOUD

Python Code

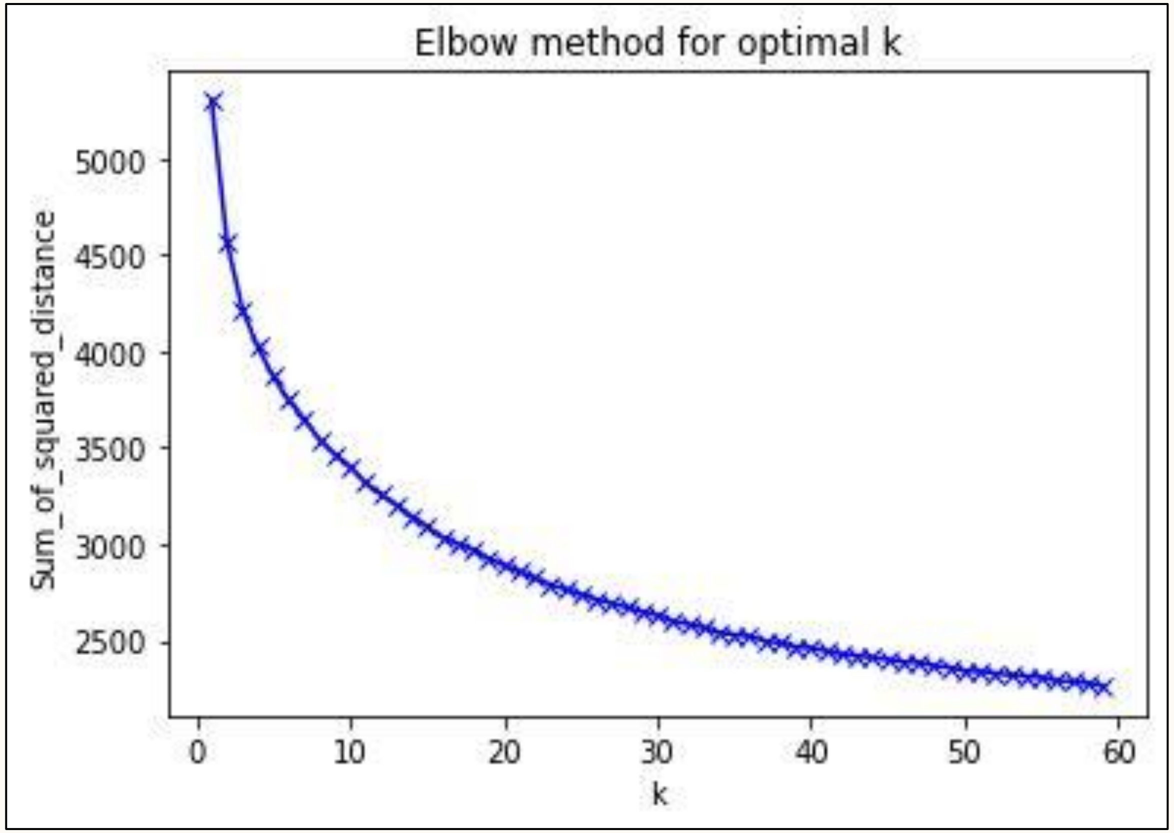
```
from wordcloud import WordCloud

text = " ".join(review for review in df['review'])
wordcloud = WordCloud(background_color="white").generate(text)
wordcloud.words_

# plot the WordCloud image
plt.figure(figsize = (10, 10), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)
plt.show()
```

```
{'phone': 1.0,
 'good': 0.5716429107276819,
 'mobile': 0.48799699924981244,
 'lenovo': 0.3522130532633158,
 'camera': 0.34433608402100524,
 'battery': 0.3420855213803451,
 'problem': 0.30195048762190546,
 'issue': 0.2873218304576144,
 'nice': 0.24118529632408103,
 'feature': 0.24006001500375093,
 'time': 0.231807951987997,
 'working': 0.23143285821455364,
 'good phone': 0.23143285821455364,
 'even': 0.22918229557389347,
 'amazon': 0.2250562640660165,......................}
```

Elbow Chart And Gap Statistic

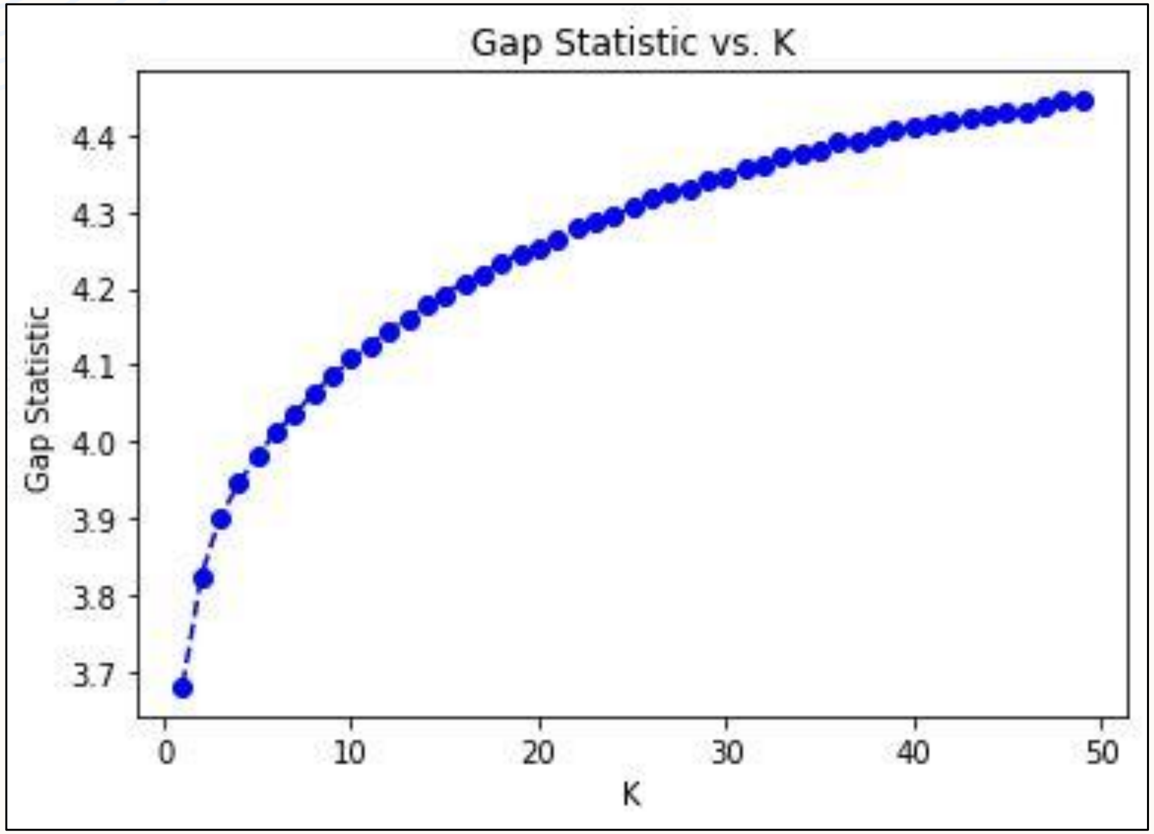



```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidfconvert =
TfidfVectorizer(analyzer=clean_text,ngram_range=(1,3)).fit(df['review'])
len(tfidfconvert.vocabulary_)
x_transformed = tfidfconvert.transform(df['review'])


from sklearn.cluster import KMeans
Sum_of_squared_distance =[]
K = range(1,60)
for k in K:
        km = KMeans(n_clusters=k)
        km = km.fit(x_transformed )
        Sum_of_squared_distance.append(km.inertia_)

plt.plot(K,Sum_of_squared_distance,'bx-')
plt.xlabel('k')
plt.ylabel('Sum_of_squared_distance')
plt.title('Elbow method for optimal k')
plt.show()
```



Based on the elbow chart and gap statistic :
Prefer to do the topic modelling using K=10

CREATING DICTIONARY


```
#lemmatization

import spacy
nlp = spacy.load('en_core_web_md')

def lemmatization(texts,allowed_postags=['NOUN', 'ADJ']):
      output = []
      for sent in texts:
            doc = nlp(sent)
            output.append([token.lemma_ for token in doc if token.pos_ in
allowed_postags ])
      return output

text_list=df['review'].tolist()
print(text_list[1])
tokenized_reviews = lemmatization(text_list)
print(tokenized_reviews[1])

#Create vocabulary dictionary and document term matrix

dictionary = corpora.Dictionary(tokenized_reviews)
doc_term_matrix = [dictionary.doc2bow(rev) for rev in tokenized_reviews]
```

Length of dictionary =9371

Dictionary(9371 unique tokens: ['good', 'improvement', 'need', 'backup', 'bad']...)

TOPIC MODELLING - LDA

```
import gensim
from gensim import corpora

# Creating the object for LDA model using gensim library

LDA = gensim.models.ldamodel.LdaModel

# Build LDA model

lda_model = LDA(corpus=doc_term_matrix,
id2word=dictionary, num_topics=6, random_state=100,
                chunksize=1000, passes=50,iterations=100)
```

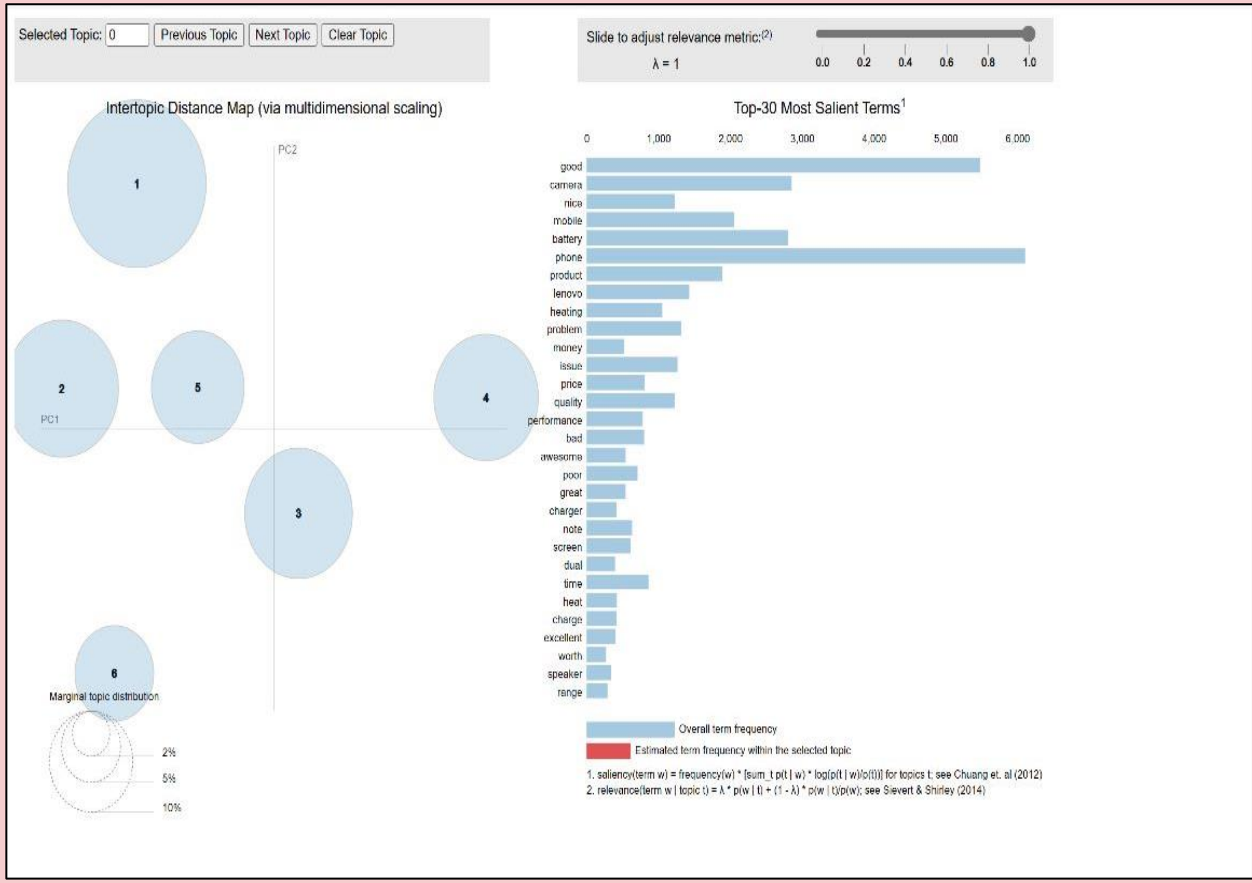
TOPIC MODELLING - LDA

Result:

Topic 0	mobile, product, money, bad, nice, delivery, value, waste, amazon, worth
Topic 1	problem, heating, phone, issue, battery, heat, excellent, awesome, drain, network
Topic 2	poor, phone, issue, network, lenovo, video, range, time, quality, battery
Topic 3	phone, battery, time, charge, charger, hour, turbo, full, day, usage
Topic 4	good, battery, product, backup, performance, life, smartphone, thank, awesome, quality
Topic 5	camera, quality, sound, speaker, mode, average, music, dolby, depth, happy
Topic 6	update, software, working, worth, super, processor, oreo, experience, love, volte
Topic 7	camera, good, phone, price, quality, great, performance, dual, feature, overall
Topic 8	phone, nice, service, lenovo, screen, bad, month, glass, customer, day
Topic 9	lenovo, note, device, call, screen, many, option, problem, feature, available

TOPIC MODELLING - LDA

Result:



TOPIC MODELLING - LDA

Result:

Topic: 0 keywords are

nice, mobile, product, money, phone, worth, delivery, value, glass, waste

Topic: 1 keywords are

battery, phone, heating, problem, issue, bad, product, mobile, awesome, time

Topic: 2 keywords are

lenovo, phone, poor, note, issue, time, call, mobile, update, battery

Topic: 3 keywords are

phone, price, great, charger, screen, range, service, month, feature, turbo

Topic: 4 keywords are

good, phone, camera, battery, quality, performance, product, price, backup, life

Topic: 5 keywords are

camera, quality, dual, speaker, mode, front, sound, average, amazing, depth

Perplexity: -7.9740844536125355

Coherence Score: 0.55015485303433

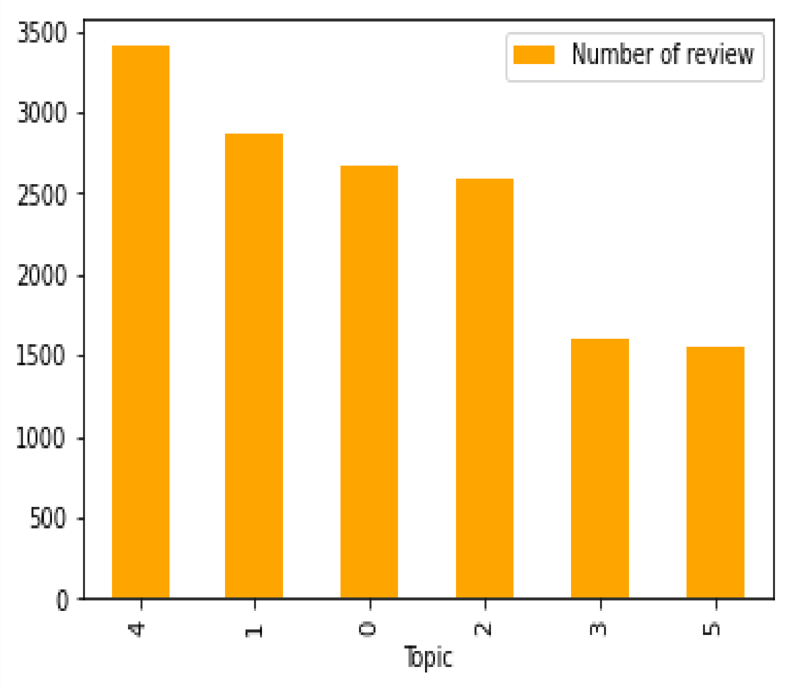
TOPIC MODELLING - LDA

Result:

Topic Number	Key word	Topic Label
Topic 0	nice, mobile, product, money, phone, worth, delivery, value, glass, waste	**Value for Money**
Topic 1	battery, phone, heating, problem, issue, bad, product, mobile, awesome, time	**Battery**
Topic 2	lenovo, phone, poor, note, issue, time, call, mobile, update, battery	**Software**
Topic 3	phone, price, great, charger, screen, range, service, month, feature, turbo	**Design**
Topic 4	good, phone, camera, battery, quality, performance, product, price, backup, life	**Quality**
Topic 5	camera, quality, dual, speaker, mode, front, sound, average, amazing, depth	**Performance**

DOMINANT TOPIC

Topic Number	Topic Label	Number of review
0	**Value for Money**	2669
1	**Battery**	2864
2	**Software**	2582
3	**Design**	1596
4	**Quality**	3406
5	**Performance**	1558



Topic 0:	
Positive review	1664
Negative review	1005

Topic 1:	
Positive review	2002
Negative review	842

Topic 2:	
Positive review	1938
Negative review	644

Topic 3:	
Positive review	801
Negative review	795

Topic 4:	
Positive review	2431
Negative review	975

Topic 5:	
Positive review	977
Negative review	581

TEXT CLASSIFICATION

FRAME WORK

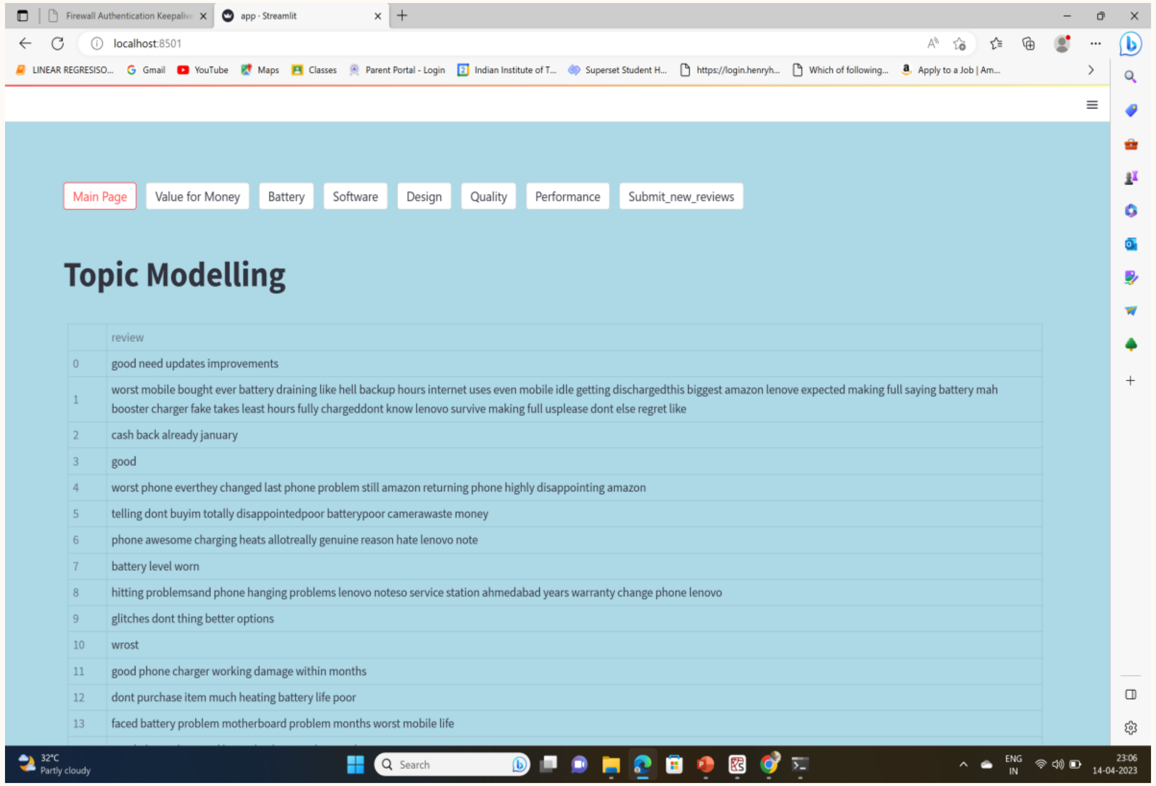
Review → Data cleaning → Tfidf vector → Model training

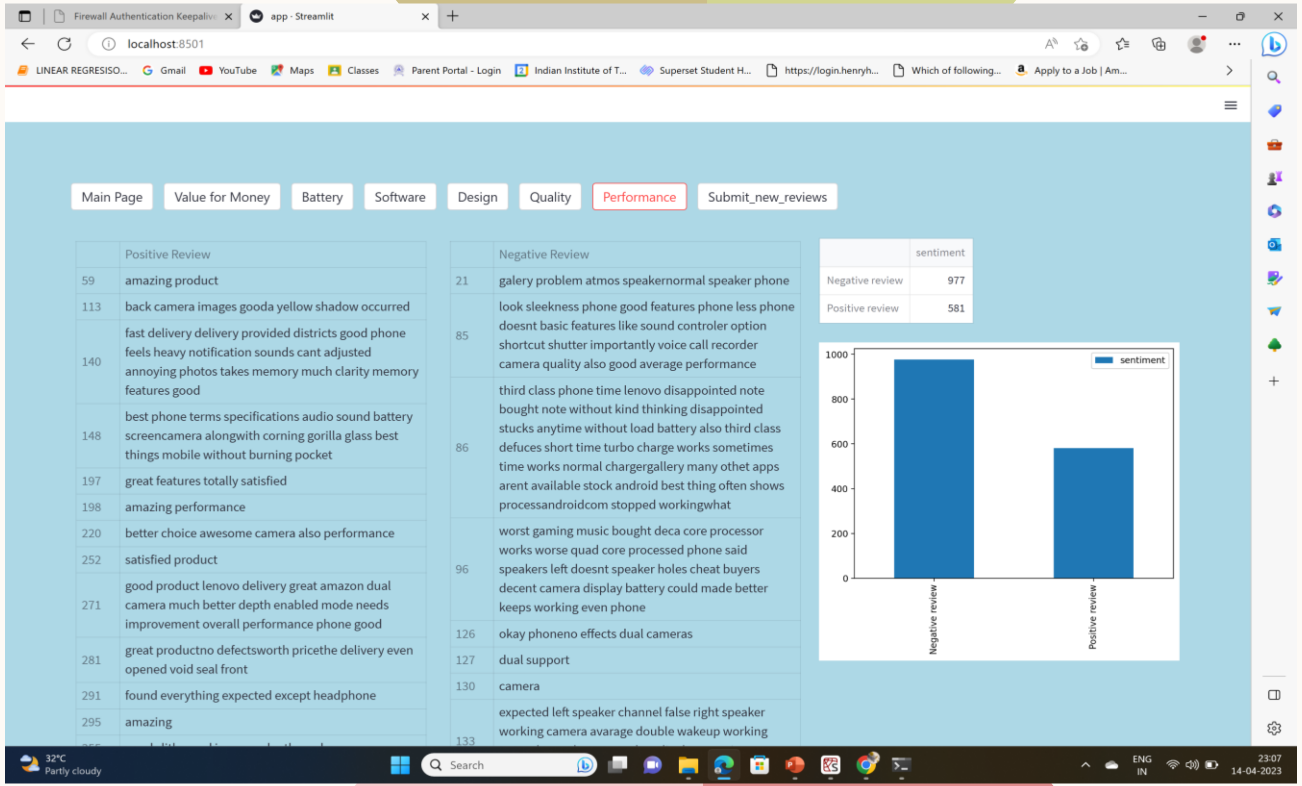
RESULT – RANDOM FOREST CLASSIFIER

		Predicted +	Predicted -
Actual	+	**1342**	**177**
Actual	-	322	1094

	precision	recall	f1-score	support
0	0.81	0.88	0.84	1519
1	0.86	0.77	0.81	1416
accuracy			0.83	2935
macro avg	0.83	0.83	0.83	2935
weighted avg	0.83	0.83	0.83	2935

WEB PAGE


localhost:8501
Main Page
Value for Money
Battery
Software
Design
Quality
Performance
Submit_new_reviews
Topic Modelling
review
0 good need updates improvements
1 worst mobile bought ever battery draining like hell backup hours internet uses even mobile idle getting dischargedthis biggest amazon lenove expected making full saying battery mah booster charger fake takes least hours fully chargeddont know lenovo survive making full usplease dont else regret like
2 cash back already january
3 good
4 worst phone everthey changed last phone problem still amazon returning phone highly disappointing amazon
5 telling dont buyim totally disappointedpoor batterypoor camerawaste money
6 phone awesome charging heats allotreally genuine reason hate lenovo note
7 battery level worn
8 hitting problemsand phone hanging problems lenovo noteso service station ahmedabad years warranty change phone lenovo
9 glitches dont thing better options
10 wrost
11 good phone charger working damage within months
12 dont purchase item much heating battery life poor
13 faced battery problem motherboard problem months worst mobile life


localhost:8501
Main Page
Value for Money
Battery
Software
Design
Quality
Performance
Submit_new_reviews
Positive Review
59 amazing product
113 back camera images gooda yellow shadow occurred
140 fast delivery delivery provided districts good phone feels heavy notification sounds cant adjusted annoying photos takes memory much clarity memory features good
148 best phone terms specifications audio sound battery screencamera alongwith corning gorilla glass best things mobile without burning pocket
197 great features totally satisfied
198 amazing performance
220 better choice awesome camera also performance
252 satisfied product
271 good product lenovo delivery great amazon dual camera much better depth enabled mode needs improvement overall performance phone good
281 great productno defectsworth pricethe delivery even opened void seal front
291 found everything expected except headphone
295 amazing
Negative Review
21 galery problem atmos speakernormal speaker phone
85 look sleekness phone good features phone less phone doesnt basic features like sound controler option shortcut shutter importantly voice call recorder camera quality also good average performance
86 third class phone time lenovo disappointed note bought note without kind thinking disappointed stucks anytime without load battery also third class defuces short time turbo charge works sometimes time works normal chargergallery many othet apps arent available stock android best thing often shows processandroidcom stopped workingwhat
96 worst gaming music bought deca core processor works worse quad core processed phone said speakers left doesnt speaker holes cheat buyers decent camera display battery could made better keeps working even phone
126 okay phoneno effects dual cameras
127 dual support
130 camera
133 expected left speaker channel false right speaker working camera avarage double wakeup working
sentiment
Negative review 977
Positive review 581
1000
800
600
400
200
0
sentiment
Negative review
Positive review

THANK YOU