

Amazon Kindle Store Reviews Analysis using IBM Watson services

1. INTRODUCTION

1.1 Overview

Amazon Kindle Store is an e-book e-commerce store for all the book reading hobbyists. Online reviews are a category of product information created by users based on personal handling experience. Online shopping websites endow with platforms for consumers to review products and carve up opinions. The problem is most of the comments from customer reviews about the products are contradicted to their ratings. Many customers will post their comments and forgot to rate the product or not engrossed to rate it.

Sentiment mining plays a very important role in business to understand the opinion of customers to improve the products. Customer also depends on the opinion of others who have bought the products already. Reviews or feedback becomes the deciding factor to buy or sell a product. A rating of the products gives a speedy clarification to pact with the product. We will be using Natural language processing to analyse the sentiment (positive or a negative) of the given review.

1.2 Purpose

Sentiment mining plays a very important role in business to understand the opinion of customers to improve the products. Customer also depends on the opinion of others who have bought the products already. Reviews or feedback becomes the deciding factor to buy or sell a product. A rating of the products gives a speedy clarification to pact with the product. We will be using Natural language processing to analyse the sentiment (positive or a negative) of the given review.

2. LITERATURE SURVEY

2.1 Existing problem

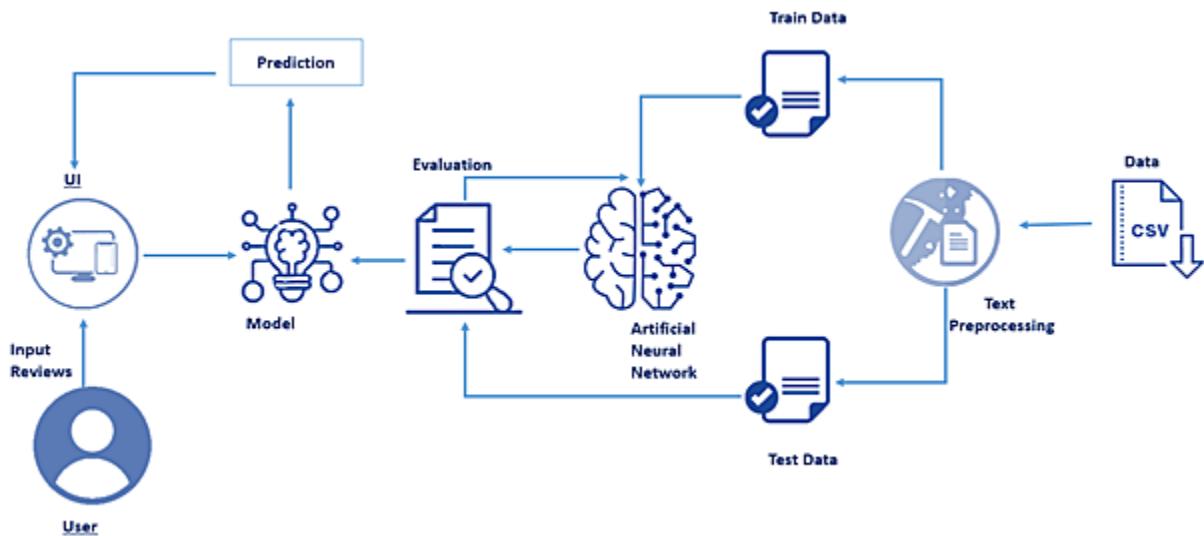
As the commercial sites of the world are almost fully online platforms, people are trading products through different e-commerce websites. And for that reason reviewing products before buying is also a common scenario. Also nowadays, customers are more inclined towards the reviews to buy a product. So analyzing the data from those customer reviews to make the data more dynamic is an essential field nowadays. In this age of increasing machine learning and deep learning based algorithms, reading thousands of reviews to understand a product is rather time consuming where we can polarize a review on a particular category to understand its popularity among the buyers all over.

2.2 Proposed solution

The objective of this paper is to categorize the positive and negative feedback of the customers over different products and build a supervised learning model to polarize large amounts of reviews. Any online item with a large amount of positive reviews provides a powerful comment of the legitimacy of the item. Conversely, books, or any other online item, without reviews puts potential prospects in a state of distrust. Quite simply, more reviews look more convincing. People value the consent and experience of others and the review on a material is the only way to understand others' impression on the product.

3. THEORITICAL ANALYSIS

3.1 Block Diagram



3.2 Hardware / Software designing

Software Requirements:

- Anaconda Navigator
- Keras
- Flask

Hardware Requirements:

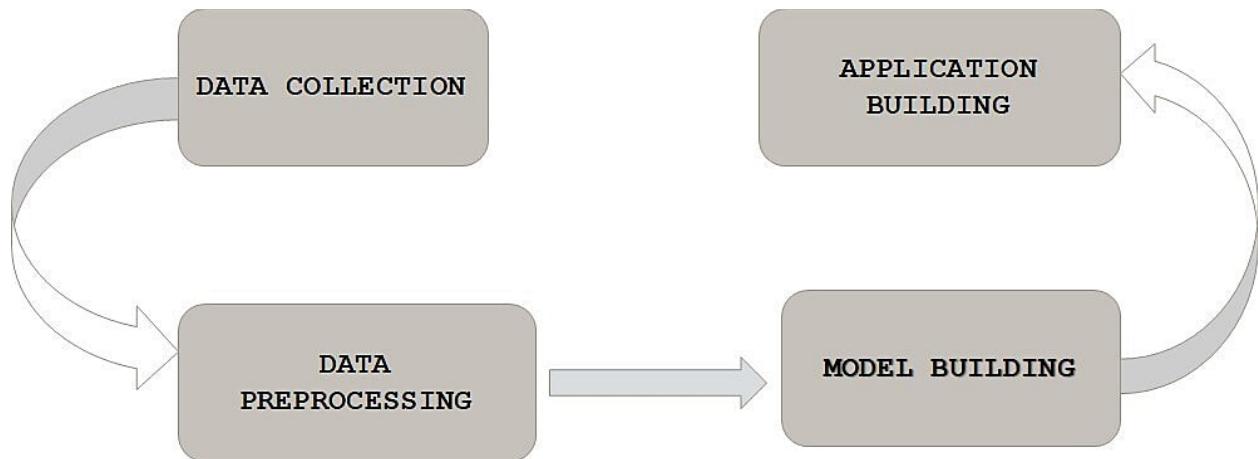
- Processor : Intel Core i3
- Hard Disk Space : Min 100 GB
- Ram : 4 GB
- Display : 14.1 "Color Monitor(LCD, CRT or LED)
- Clock Speed : 1.67 GHz

4. EXPERIMENTAL INVESTIGATIONS

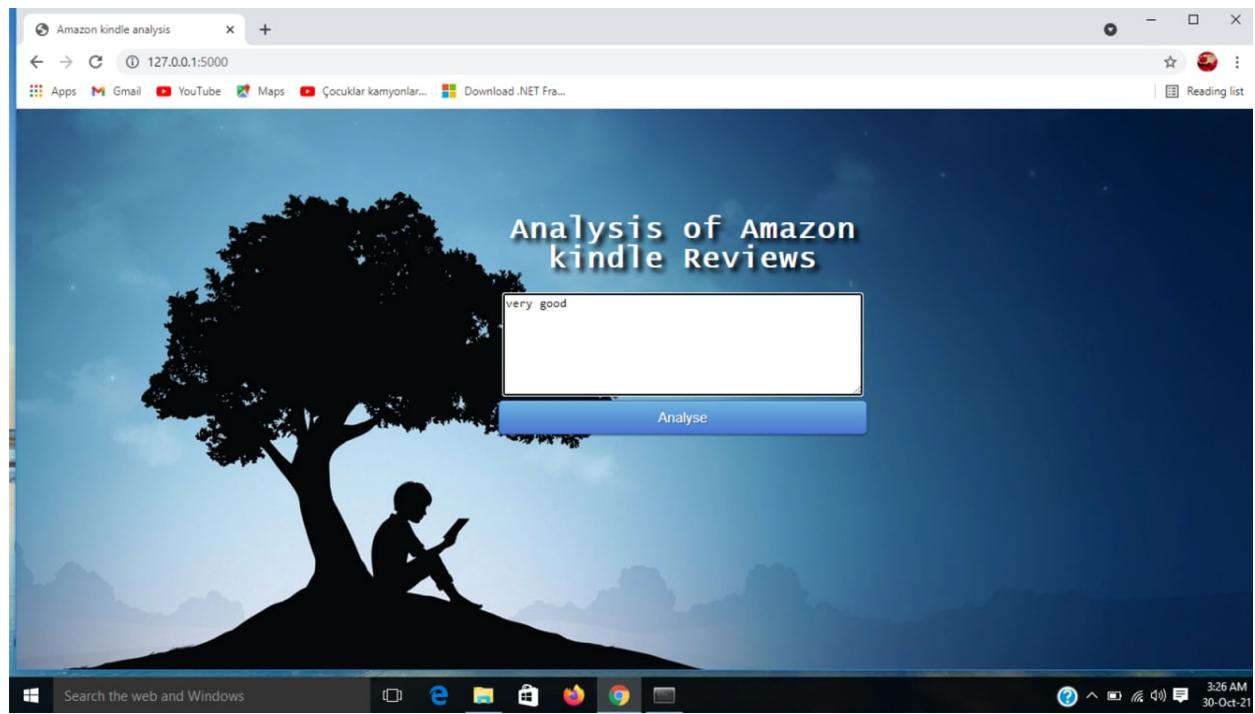
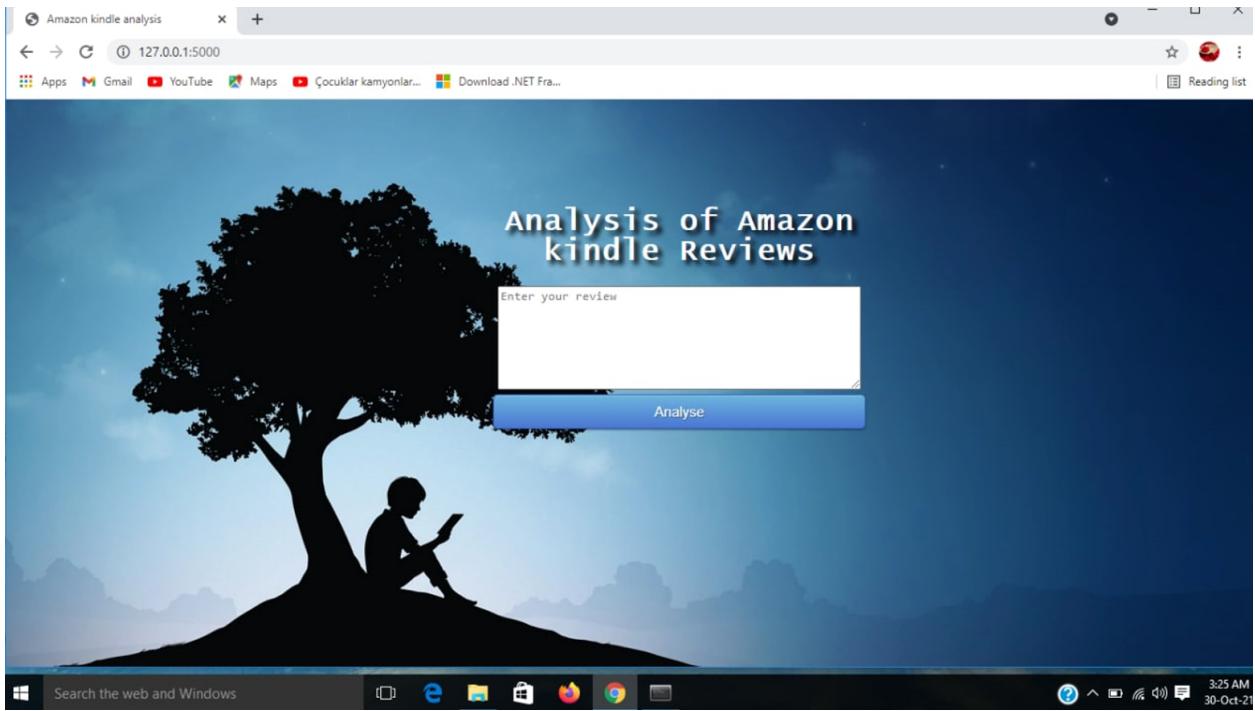
Our experimental results show that the proposed approach can aid in the accurate and timely review of the comments which is positive or negative sense.

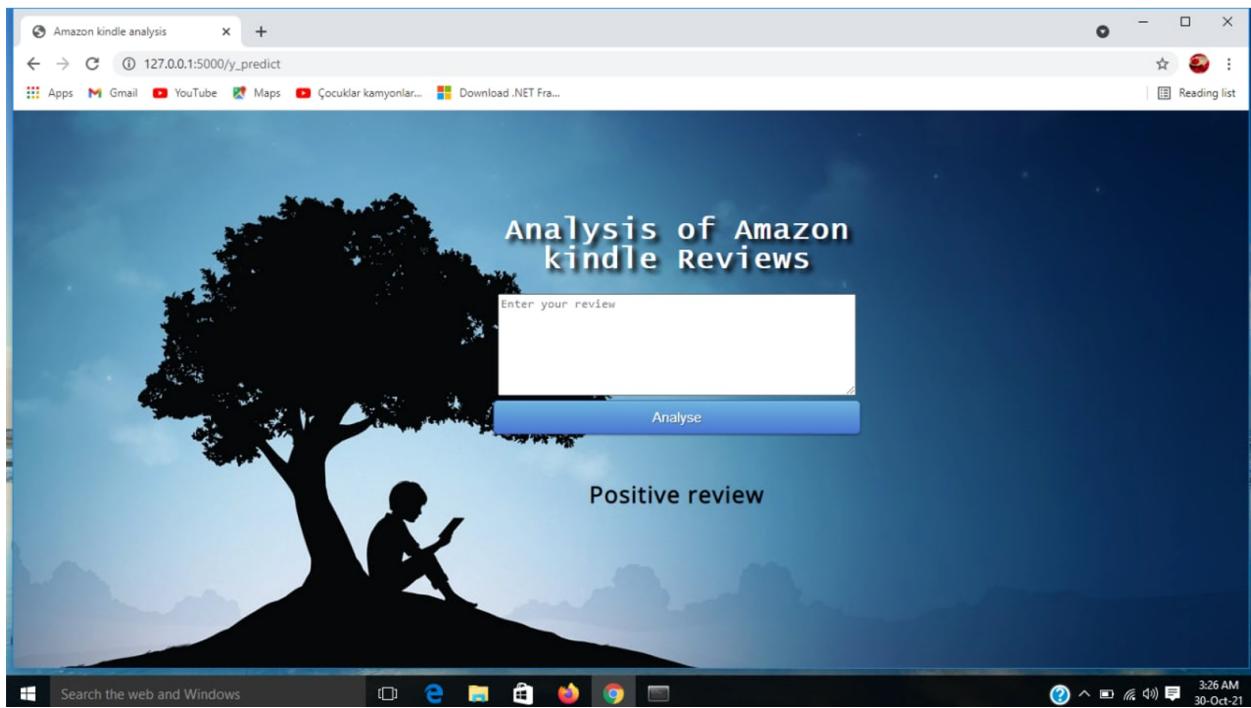
The experimental results achieved 96.51% accuracy demonstrating the effectiveness of the proposed technique for identifying positive and negative comments. The result of the analysis is carried out whether the given comment is positive or negative

5. FLOWCHART



6. RESULT





7. ADVANTAGES & DISADVANTAGES

Advantages:

Sentiment mining plays a very important role in business to understand the opinion of customers to improve the products. Customer also depends on the opinion of others who have bought the products already. Reviews or feedback becomes the deciding factor to buy or sell a product. A rating of the products gives a speedy clarification to pact with the product. We will be using Natural language processing

to analyse the sentiment (positive or a negative) of the given review. Easy way to know the comment is good or bad

8. APPLICATIONS

- The main aim of the application is comment reviewing ,this application is to provide proper review of the comment given by the user.
- This application is helpful to all the users who use the amazon kindle. It is a user friendly application

9. CONCLUSION

It is completely impossible to use only raw text as input for making predictions. Hence,we saw that the pre-processing step played a major role in the complete process of NLP. To get better results, accuracy and make the machine take all the text as tokens, pre-processing of data is to be done carefully looking at the type of contents present in it. The most important thing is to be able to extract the relevant features from the given source of data. This kind of data can often come as a good complementary source in order to extract more learning features and increase the predictive power of the models. And the user is able to predict that the given comment is positive or negative.

10. FUTURE SCOPE

In future, the work can be extended to perform multi-class classification of reviews which will provide a delineated nature of review to the consumer, hence better judgment of the product. It can also be used to predict the rating of a product from the review. This will provide users with a reliable rating because sometimes the rating received by the product and the sentiment of the review do not provide justice to each other. The proposed extension of work will be very beneficial for the e-commerce industry as it will augment user satisfaction and trust.

11. BIBILOGRAPHY

- <https://en.wikipedia.org/wiki/Machine-learning>
- <https://www.geeksforgeeks.org/introduction-machine-learning>
- <https://en.wikipedia.org/wiki/sentiment-mining>

APPENDIX

Source Code

The screenshot shows a Jupyter Notebook interface running on a Windows desktop. The notebook has tabs for 'Home Page - Select or create a...' and 'Amazon kindle review - Jupyter'. The main area displays the following code and its output:

```
In [2]: #import required libraries
import pandas Library
import pandas as pd
import numpy
import numpy as np
import requests
import requests
import io
import io

In [3]: #import the dataset in the data variable
data=pd.read_csv('D:/Projeoct/dataset/kindle_reviews.csv')

In [4]: data.shape
Out[4]: (982619, 10)

In [5]: data.head()
Out[5]:   Unnamed: 0      asin  helpful  overall      reviewText  reviewTime  reviewerID  reviewerName  summary  unixReviewTime
0          0  B000F83SZQ  [0, 0]       5  I enjoy vintage books and movies so I enjoyed ...
1          1  B000F83SZQ  [2, 2]       4  This book is a reissue of an old one; the auth...
2          2  B000F83SZQ  [2, 2]       4  This was a fairly interesting read. It had ol...
3          3  B000F83SZQ  [1, 1]       5  I'd never read any of the Amy Brewster myster...
4          4  B000F83SZQ  [0, 1]       4  If you like period pieces - clothing, lingo, y...
```

The data frame has 10 columns: Unnamed: 0, asin, helpful, overall, reviewText, reviewTime, reviewerID, reviewerName, summary, and unixReviewTime. The first five rows of the data frame are displayed.

The taskbar at the bottom shows various pinned icons and the system status bar indicating 31°C Rain, ENG IN, 452 PM, and 12/3/2021.

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter | +

localhost:8888/notebooks/Amazon%20kindle%20review.ipynb

jupyter Amazon kindle review (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

In [2]: #import required libraries

```
#import pandas Library
import pandas as pd
import numpy
import numpy as np
import requests
import requests
import io
import io
```

In [3]: #import the dataset in the data variable

```
data=pd.read_csv("D:/Project/dataset/kindle_reviews.csv")
```

In [4]: data.shape

Out[4]: (982619, 10)

In [5]: data.head()

Out[5]:

	Unnamed: 0	asin	helpful	overall	reviewText	reviewTime	reviewerID	reviewerName	summary	unixReviewTime
0	0	B000F83S2Q	[0, 0]	5	I enjoy vintage books and movies so I enjoyed ...	05.6.2014	A1F6404F1V2G9J	Avidreader	Nice vintage story	1399248000
1	1	B000F83S2Q	[2, 2]	4	This book is a reissue of an old one; the auth...	01.6.2014	AN0N05A9LJUEQ	critters	Different...	1389866400
2	2	B000F83S2Q	[2, 2]	4	This was a fairly interesting read. It had ol...	04.4.2014	A795DMNCJLA6	dot	Oldie	1396589600
3	3	B000F83S2Q	[1, 1]	5	I'd never read any of the Amy Brewster mysteri...	02.19.2014	A1FV0SX13TWVXQ	Elaine H. Turley "Montana Songbird"	I really liked it.	1392768000
4	4	B000F83S2Q	[0, 1]	4	If you like period pieces - clothing, lingo, v...	03.19.2014	A3SP7OKDG7WBLN	Father Dowling Fan	Period Mystery	1395187200

In [6]: #assigning 50000 rows to data

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter | +

localhost:8888/notebooks/Amazon%20kindle%20review.ipynb

jupyter Amazon kindle review (autosaved)

File Edit View Insert Cell Kernel Widgets Help

In [9]: #deleting or dropping the unwanted columns from the dataset

```
del data['Unnamed: 0']
del data['asin']
del data['helpful']
del data['reviewTime']
del data['reviewerID']
del data['reviewerName']
del data['unixReviewTime']
```

In [10]: #first 10 rows of data

```
data.head(10)
```

Out[10]:

	overall	reviewText	summary
0	5	I enjoy vintage books and movies so I enjoyed ...	Nice vintage story
1	4	This book is a reissue of an old one; the auth...	Different...
2	4	This was a fairly interesting read. It had ol...	Oldie
3	5	I'd never read any of the Amy Brewster mysteri...	I really liked it.
4	4	If you like period pieces - clothing, lingo, v...	Period Mystery
5	4	A beautiful in-depth character description mak...	Review
6	4	I enjoyed this one tho I'm not sure why it's c...	Nice old fashioned story
7	4	Never heard of Amy Brewster. But I don't need ...	Enjoyable reading and reminding the old times
8	5	Darth Maul working under cloak of darkness com...	Darth Maul
9	4	This is a short story focused on Darth Maul's ...	Not bad, not exceptional

In [11]: #checking value counts

```
data.overall.value_counts()
```

Out[11]:

5	23090
4	14980
3	7013
2	2832

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter

In [6]: #assigning 50000 rows to data
data=data.head(50000)

In [7]: #checking for null values
data.isnull().any()

Out[7]:

Unnamed: 0	False
asin	False
helpful	False
overall	False
reviewText	True
reviewTime	False
reviewerID	False
reviewerName	True
summary	False
unixReviewTime	False
dtype:	bool

In [8]: data.isnull().sum()

Out[8]:

Unnamed: 0	0
asin	0
helpful	0
overall	0
reviewText	1
reviewTime	0
reviewerID	0
reviewerName	149
summary	0
unixReviewTime	0
dtype:	int64

In [9]: #deleting or dropping the unwanted columns from the dataset
del data['Unnamed: 0']
del data['asin']
del data['helpful']

Type here to search

31°C Rain 453 PM IN 12/3/2021

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter

In [12]: #check the null values
data.isna().sum()

Out[12]:

overall	0
reviewText	1
summary	0
dtype:	int64

In [13]: #joining review description and summary into one col
data['reviewText']=data['reviewText']+data['summary']

In [14]: data.drop(['summary'],axis=1,inplace=True)

In [15]: data.head()

Out[15]:

	overall	reviewText
0	5	I enjoy vintage books and movies so I enjoyed ...
1	4	This book is a reissue of an old one; the auth...
2	4	This was a fairly interesting read. It had ol...
3	5	I'd never read any of the Amy Brewster myster...
4	4	If you like period pieces - clothing, lingo, y...

In [16]: #checking for null values
data.isna().sum()

Out[16]:

overall	0
reviewText	1
dtype:	int64

In [17]: #since there is only one null value, replace it with blank space
data['reviewText'].fillna("",inplace = True)

Type here to search

31°C Rain 453 PM IN 12/3/2021

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter | +

jupyter Amazon kindle review (autosaved)

File Edit View Insert Cell Kernel Widgets Help

In [18]: #Grouping the overall rating of scale 1-5 to 2 categories
def review_sentiment(rating):
 #(positive) and with 1(negative)
 if(rating == 5 or rating == 4 or rating==3):
 return 0
 else:
 return 1

In [19]: data.overall = data.overall.apply(review_sentiment)

In [20]: data.overall.value_counts()

Out[20]: 0 45083
1 4917
Name: overall, dtype: int64

In [21]: data.head(50)

Out[21]:

	overall	reviewText
0	0	I enjoy vintage books and movies so I enjoyed ...
1	0	This book is a reissue of an old one; the auth...
2	0	This was a fairly interesting read. It had ol...
3	0	I'd never read any of the Amy Brewster mysteri...
4	0	If you like period pieces - clothing, lingo, y...
5	0	A beautiful in-depth character description mak...
6	0	I enjoyed this one tho I'm not sure why it's c...
7	0	Never heard of Amy Brewster. But I don't need ...
8	0	Darth Maul working under cloak of darkness com...
9	0	This is a short story focused on Darth Maul's ...

Type here to search

31°C Rain ENG IN 453 PM 12/3/2021

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter | +

jupyter Amazon kindle review (autosaved)

File Edit View Insert Cell Kernel Widgets Help

In [22]: len(list(data["overall1"]))

Out[22]: 50000

In [23]: #split the data into x(independent variable)
x=data.iloc[:,1].values

In [24]: #import natural language toolkit
#import nltk
#nltk.download("stopwords")
#nltk.download("wordnet")
#import stopwords library to remove stopwords
from nltk.corpus import stopwords
#library used for stem the words
from nltk.stem.porter import PorterStemmer
#create an object for stemming
ps = PorterStemmer()
#library used for stem the words
from nltk.stem import WordNetLemmatizer
#create an object for wordnet lemmatizer
wordnet=WordNetLemmatizer()

In [25]: import re

In [26]: # Initialize empty array to append clean text
corpus=[]
no of rows to clean
for i in range(len(x)):
 #replacing punctuations and numbers using re Library
 temp=re.sub('[^a-zA-Z]',' ',x[i])
 # convert all text to lower cases
 temp=temp.lower()
 # split to array(default delimiter is " ")
 temp=temp.split()

Type here to search

31°C Rain ENG IN 454 PM 12/3/2021

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter | +

jupyter Amazon kindle review (autosaved)

File Edit View Insert Cell Kernel Widgets Help

In [31]: #checking x_train shape
x_train.shape
Out[31]: (40000,)

In [32]: #checking y_train shape
y_train.shape
Out[32]: (40000,)

In [33]: #checking x_test shape
x_test.shape
Out[33]: (10000,)

ANN

```
In [ ]: #import libraries
import tensorflow
import keras
import Sequential
from tensorflow.keras.models import Sequential
import Dense
from tensorflow.keras.layers import Dense
```

```
In [ ]: #initialize the model
model=Sequential()

#adding input layer
model.add(Dense(input_dim=6000,kernel_initializer='random_uniform',
activation='relu',units=160))

#adding hidden layer
```

Type here to search

31°C Rain ENG IN 454 PM 12/3/2021

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter | +

jupyter Amazon kindle review (autosaved)

File Edit View Insert Cell Kernel Widgets Help

```
#configure the learning process
model.compile(optimizer='adam',loss='binary_crossentropy',
metrics=[ 'accuracy'])

#training the model
model.fit(x_train,y_train,epochs=20,batch_size=32)
```

```
In [ ]: #save the model
model.save("amazo.h5")
```

```
In [ ]: ypred=model.predict(x_test)
```

```
In [ ]: ypred
```

```
In [ ]: #save bag of word model
import joblib
joblib.dump(cv.vocabulary_, "amazo.save")
```

```
In [ ]: loaded=CountVectorizer(decode_error='replace',vocabulary=joblib.load('amazo.save'))
```

```
In [ ]: d="Writing was good"
d=d.split('delimiter')
result=model.predict(loaded.transform(d))
print(result)
prediction=result>0.5
#print(prediction)
if prediction[0] == False:
    print("Positive review")
else if prediction[0] == True:
    print("Negative review")
```

```
In [ ]: from tensorflow.keras.models import load_model
model=tf.keras.models.load_model("amazo.h5")
```

Type here to search

31°C Rain ENG IN 454 PM 12/3/2021

Home Page - Select or create a new notebook | Amazon kindle review - Jupyter Notebook

jupyter Amazon kindle review (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3

```
print("Negative review")
```

```
In [ ]: from tensorflow.keras.models import load_model  
model=tf.keras.models.load_model("amazo.h5")
```

```
In [ ]: #import Load model function  
from tensorflow.keras.models import load_model  
#load our saved model file  
model=tf.keras.models.load_model("amazo.h5")  
#import countvectorizer  
from sklearn.feature_extraction.text import CountVectorizer  
import joblib  
#load saved bag of word model file  
loaded=CountVectorizer(decode_error='replace',vocabulary=joblib.load('amazo.save'))
```

```
In [ ]: d="Writing was good"  
d=d.split('delimiter')  
result=model.predict(loaded.transform(d))  
print(result)  
prediction=result>0.5  
#print(prediction)  
if prediction[0] == False:  
    print("Positive review")  
elif prediction[0] == True:  
    print("Negative review")
```

```
In [ ]:
```

Windows Start Type here to search 31°C Rain ENG 454 PM IN 12/3/2021