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

A book review helps other users get a clear idea of the book before reading it. They can read the reviews and make their mind clear, and decide whether the book is worth read or not.

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.

2.LITERATURE SURVEY

2.1 Existing system

In the existing system it is not easy to identify the review is correct or not. And also we want to ensure that the review is a valid statement.

It is very time consuming task that to validate the review is true.

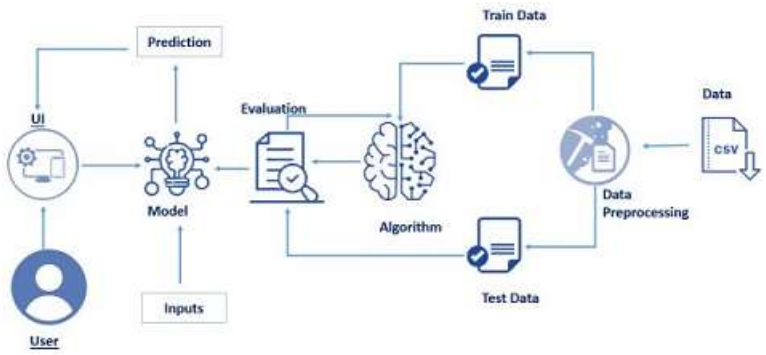
2.2 Proposed system

Amazon Kindle is a series of e-readers designed and marketed by Amazon. Amazon Kindle devices enable users to browse, buy, download, and read e-books, newspapers, magazines and other digital media via wireless networking to the Kindle Store.

Using deep learning algorithms we have to check whether the review is positive or negative. And also it is easy to identify the review about the kindle store books.

3.THEORETICAL ANALYSIS

3.1 Block diagram



3.2 Software/ Hardware designing

Hardware Requirements:
Processor : Intel Core I3
RAM : 8.00 GB
Storage 250 GB
OS : Windows/Linux/MAC

Software Requirements:

Operating System: Windows 10 Home

Anaconda:Anaconda must be installed as it provides jupyter notebook and Spyder

IBM Academic initiative account is required to access IBM Services

IBM Watson Studio IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud. IBM Watson Machine Learning - IBM Watson Machine Learning helps data scientists and developers accelerate Al and machine-learning deployment. IBM Cloud Object Storage IBM Cloud Object Storage makes it possible

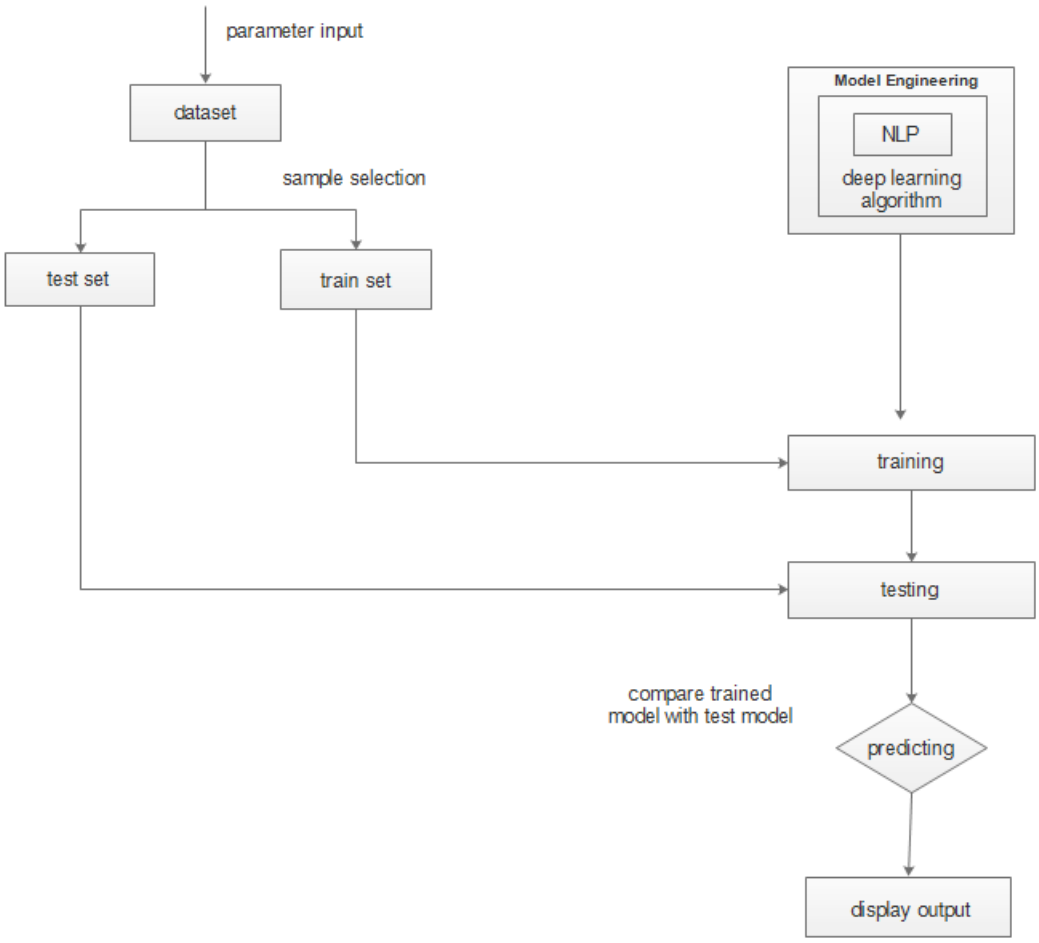
4. EXPERIMENTAL INVESTIGATION

We will be creating and testing our model for predicting if a review is negative or positive. since there are multiple algorithms, we can use to build our model, we will compare the accuracy scores after testing and pick the most accurate algorithm.

From the list, we are using Natural language processing which gives the highest accuracy and select it as our algorithm of choice for future use.

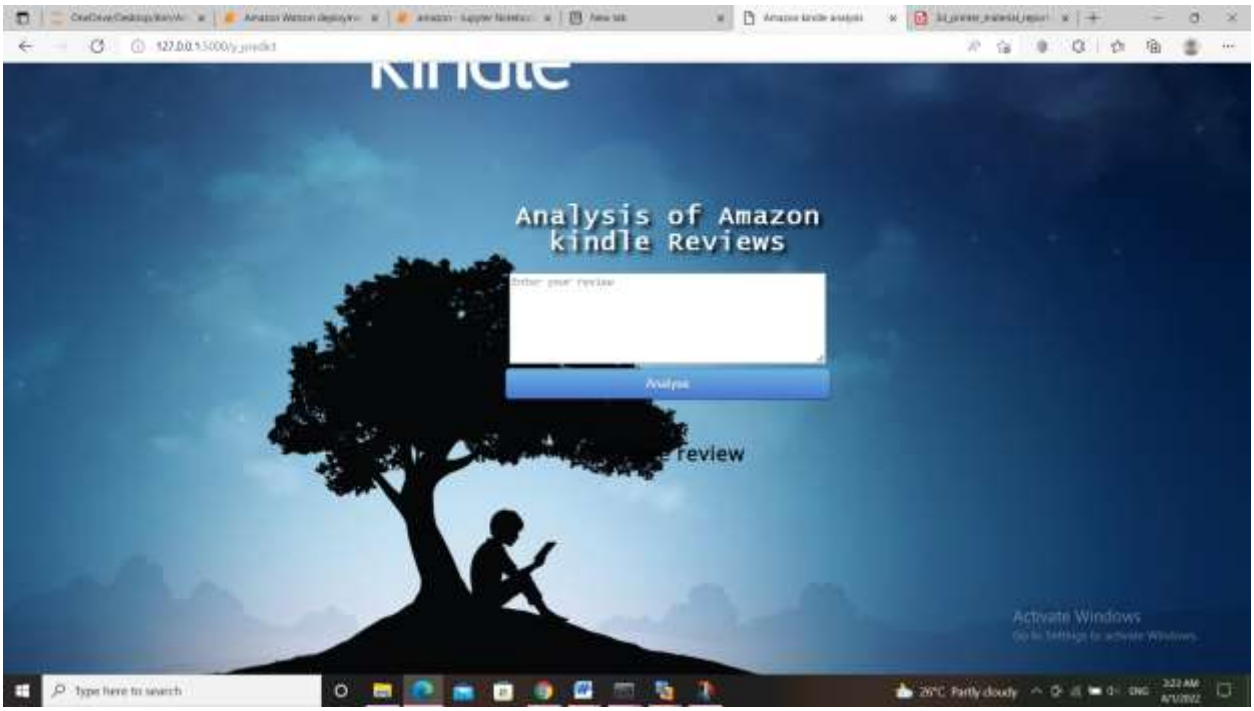
On the basis of results we have done the conclusion the NLP is the most accurate model which we have tested.

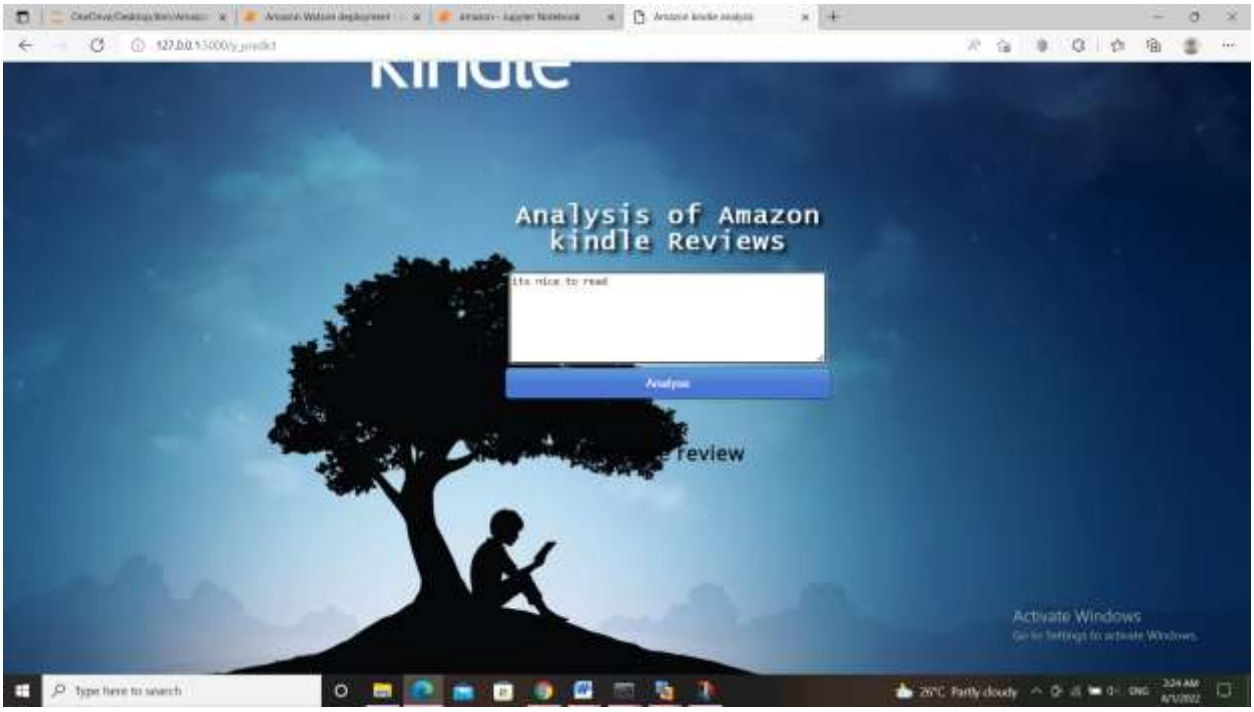
5.FLOWCHART



6.RESULT

The final result of the Amazon Kindle review store is using NLP predict the review is positive or negative.





7. ADVANTAGES AND DISADVANTAGES

7.1 Advantages

The model is fast and accurate and it gives the exact reviews about kindle store books. Its very time consuming because we use NLP in this project. We get the output easily when we analyse so user can't wait much time for the output. It is very useful for a new customer who hasn't any idea about the book or product.

7.2 Disadvantages

A lack of negative review will affect the downloading of a book. So it's a big disadvantage. We use deep learning to implement this system so its need a large amount of data to predict the book reviews. The review is depend only on the interest of a customer so it may vary depend on the persons.

8.APPLICATIONS

1. From Authors side, they can know the interest of the customers.

2. From Customers side, they can know the review as negative or positive.

3. Kindle store became more user friendly.

9.CONCLUSION

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.

10. FUTURE SCOPE

Deep learning and machine learning are the growing technologies so we can easily update the system in future. It is very useful for the new users so there is always a future scope for the system.

11. BIBILIOGRAPHY

1)https://machinelearningmastery.com/

2)http://www.kaggle.com/

3)https://www.geeksforgeeks.org/decision-tree/

4)https://arxiv.org/

5)https://cloud.ibm.com/docs/create-deploy-retrain-machine-learning-mode

SOURCE CODE

In [2]:
```
from sklearn.feature_extraction.text import CountVectorizer
```

In [3]:
```
import tensorflow as tf
from tensorflow.keras import backend as k
```

In [4]:
```
import pandas as pd
import numpy as np
import  nltk #natural language tool kit
import re #regular expression -removing the special characters
from  nltk.stem.porter import PorterStemmer
from nltk.corpus import stopwords
from sklearn.model_selection import train_test_split
```

In [5]:
```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
```

In [6]:
```
from tensorflow.keras import models
```

In [ ]:
```

```

In [7]:
```
#import required libraries

#import pandas library
import pandas as pd
#import numpy
import numpy as np
#import requests
import requests
#import io
import io
import os
```

In [9]:
```
#import the dataset in the data variable
data=pd.read_csv("kindle_reviews.csv")
```

In [10]:
```
data.shape
```

Out[10]:
```
(982619, 10)
```

In [11]:
```
data.head()
```

Out[11]:

	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 ...	05 5, 2014	A1F6404F1VG29J	Avidreader	Nice vintage story	1399248000
1	1	B000F83SZQ	[2, 2]	4	This book is a reissue of an old one; the auth...	01 6, 2014	AN0N05A9LIJEQ	critters	Different...	1388966400
2	2	B000F83SZQ	[2, 2]	4	This was a fairly interesting read. It had ol...	04 4, 2014	A795DMNCJILA6	dot	Olde	1396569600
3	3	B000F83SZQ	[1, 1]	5	I'd never read any of the Amy Brewster myster...	02 19, 2014	A1FV0SX13TWVXQ	Elaine H. Turley "Montana Songbird"	I really liked it.	1392768000
4	4	B000F83SZQ	[0, 1]	4	If you like period pieces - clothing, lingo, y...	03 19, 2014	A3SPTOKDG7WBLN	Father Dowling Fan	Period Mystery	1395187200

In [12]:
```
#assigning 50000 rows to data
data=data.head(50000)
```

In [13]:
```
#checking for null values
data.isnull().any()
```

Out[13]:
```
Unnamed: 0        False
asin              False
helpful           False
overall           False
reviewText         True
reviewTime        False
reviewerID        False
reviewerName       True
summary           False
unixReviewTime    False
dtype: bool
```

In [14]:
```
data.isnull().sum()
```

Out[14]:
```
Unnamed: 0          0
asin                0
helpful             0
overall             0
reviewText          1
reviewTime          0
reviewerID          0
reviewerName      149
summary             0
unixReviewTime      0
dtype: int64
```

In [15]:
```
#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']
```


In [17]:
```
#checking value counts
data.overall.value_counts()
```

Out[17]:
```
5    23090
4    14980
3     7013
2     2832
1     2085
Name: overall, dtype: int64
```

In [18]:
```
#check the null values
data.isna().sum()
```

Out[18]:
```
overall       0
reviewText    1
summary       0
dtype: int64
```

In [19]:
```
#joining review description and summary into one col
data['reviewText']=data['reviewText']+" "+data['summary']
```

In [20]:
```
data.head()
```

Out[20]:

	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, y...	Period Mystery

In [21]:
```
data.drop(['summary'],axis=1,inplace=True)
```

In [22]:
```
#checking for null values
data.isna().sum()
```

Out[22]:
```
overall       0
reviewText    1
dtype: int64
```

In [23]:
```
#since there is only one null value, replace it with blank space
data['reviewText'].fillna("",inplace = True)
```

In [24]:
```
#Grouping the overall rating of scale 1-5 to 2 categories
def review_sentiment(rating):
    #0(positive) and  with 1(negative)
    if(rating == 5 or rating == 4 or rating==3):
        return 0
    else:
        return 1
```

In [25]:
```
data.overall = data.overall.apply(review_sentiment)
```

In [26]:
```
data.overall.value_counts()
```

Out[26]:
```
0    45083
1     4917
Name: overall, dtype: int64
```