SENTIMENT ANALYSIS OF REAL TIME PRODUCT REVIEWS

A Project Report
Submitted by

ROMIT CHANGANI (E004)
PRATYAKSH JAIN (E017)
TANISH KORGAONKAR (E021)
KARTHIK RAM SRINIVAS (E055)

Under the Guidance of

DR. PRAVIN SHRINATH

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

COMPUTER SCIENCE AND BUSINESS SYSTEMS

At

MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT AND ENGINEERING

MARCH 2023

DECLARATION

We, Romit Changani, Pratyaksh Jain, Tanish Korgaonkar and Karthik Ram Srinivas, Roll No. E004, E017, E021, E055 B. Tech(Computer Science and Business Systems), VII-VIII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs, or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages' paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source: IEEE, The institute, Dec. 2004)

4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism, and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be able to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Students:

Name: Romit Changani, Pratyaksh Jain, Tanish Korgaonkar, Karthik Ram Srinivas

Roll No.: E004, E017, E021, E055

Place: Mumbai

Date: 25/03/23

CERTIFICATE

This is to certify that the project entitled “Sentiment Analysis of Real Time Product Reviews” is the bonafide work carried out by Romit Changani, Pratyaksh Jain, Tanish Korgaonkar and Karthik Ram Srinivas, of B. Tech (Computer Science and Business Systems), MPSTME (NMIMS), Mumbai, during the VIIth and VIIIth semester of the academic year 2022-2023, in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

_____

Dr. Pravin Shrinath

Internal Mentor

_____ _____

Examiner 1 Examiner 2

_____

HOD

Abstract

Sentiment analysis is contextual mining of text which identifies and extracts subjective information in text, and helps a business to understand the social sentiment of their brand, product or service while monitoring online reviews. The reviews provided by the customers on e- commerce websites not only helps the owners to improve their services accordingly but also helps other customers to get an opinion of the products before buying them online. Our project was aimed at developing a web application to perform sentiment analysis on amazon product reviews by leveraging machine learning algorithms. We used SVM, Random Forest and Logistic Regression to train our model from which Logistic regression performed the best. We also implemented a web scraper using Beautiful Soup that gets real time reviews from amazon. The ML model was then used to predict the sentiment of each review which was displayed using interactive visualizations. With online shopping becoming the new trend this application is a one stop solution for business to track the popularity and sentiment of their product.

Table of Contents

List of Figures

Abbreviations

Abbreviations	Description
CSS	Cascading Style Sheets
HTML	Hyper Text Markup Language
NB	Naïve Bayes
NLTK	Natural Language Toolkit
SVM	Support Vector Machine
TFIDF	Term Frequency – Inverse Document Frequency
UML	Unified Modelling Language
URL	Uniform Resource Locator

Chapter 1

Introduction

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis (or opinion mining) uses Natural Language Processing to determine whether data is positive, negative or neutral. Sentiment analysis invokes to the study of text analysis, natural language processing, computational linguistic to scientifically identify, extract and study subjective information from the textual data. General meaning of sentiment analysis is to determine the insolence of a speaker, writer, or other subject with respect to particular topic or contextual polarity to a specific event, discussion, forum, interaction or any documents, etc. Due to increase in user of Internet every user is interested to put his opinion on the internet through different medium and this results opinioned data has generated on the internet. Sentiment analysis helps to analyze these opinioned data and extract some important insights which will help to other user to make decision. There are 4 types of sentiment analysis: -

Fine-grained Sentiment: This is one of the simplest and common ways of understanding the customers' sentiments. This analysis gives us an understanding of the feedback received from customers. While analyzing the sentiments, readily available categories like positive, neutral and negative are used. Providing a rating option from 1 to 5 is another way to scale the feedback given by your customers. Most e-commerce sites use this technique to know the sentiments of their customers.

Emotion Detection Sentiment Analysis: This is a more refined method of detecting the feeling in a piece of text. This kind of analysis aids to detect and understand the emotions of the people. Emotions like anger, sadness, happiness, frustration, fear, panic, worry may all be included. The upside of utilizing this is that an organization can also understand why a customer feels a specific way, but understanding the sentiments of people using emotion detection is very difficult as people use a collection of words having a different sense of meaning such as sarcasm.

Aspect-based Analysis Sentiment: This type of analysis is more focused on the aspects of a particular product or service. Aspect based sentiment analysis is essential as it can support organizations in automatically sorting and analyzing customer data, automating the processes like

customer support tasks allows us to gain significant insights on the fly. Aspect-based sentiment analysis empowers organizations to zero in on the parts of their products or administrations that their clients are griping about and helps them in fixing those issues progressively. Complaints such as glitches or major bugs in some new software applications can also be addressed.

Intent-based Sentiment Analysis: Intent classification refers to the automatic classification of textual data which is based on the customer's aim. An intent classifier can naturally dissect the writings and reports and classify them into intents like Purchase, Downgrade, unsubscribe etc. This proves helpful to comprehend the intentions behind a large number of the client's questions, automates measures, and acquires significant experiences. Intent classification empowers organizations to be more customer-friendly, especially when dealing with areas such as customer support and sales. It helps them in reacting to leads faster and handling large volumes of inquiries.

Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs. It can also be used for comparing product reviews with the competitors. The problem is to categorize the text into one specific sentiment polarity, positive or negative (or neutral) using different machine learning algorithms and to determine which of these techniques yields the best result in terms of accuracy, precision and recall. The goal is to develop a web-based application that will be able to scrape online reviews and make accurate predictions using machine learning models.

Chapter 2

Review of Literature

Sentiment analysis is a technique used to extract and measure the subjective information contained in text data. Sentiment analysis of real-time product reviews is a critical tool for businesses to understand their customers' opinions and emotions regarding their products and services. Real-time product reviews are particularly important as they provide immediate feedback to businesses, allowing them to quickly respond to any issues or concerns that customers may have.

There are several techniques available for sentiment analysis of real-time product reviews, including machine learning, deep learning, and rule-based approaches. Machine learning algorithms are widely used in sentiment analysis, as they can learn from data and improve accuracy over time. Deep learning techniques, such as deep neural networks, have also shown promising results in sentiment analysis, especially when dealing with complex and ambiguous text.

However, there are challenges in sentiment analysis of real-time product reviews. One of the primary challenges is the large volume of data that businesses must analyze. With the increasing popularity of e-commerce and social media platforms, there is an abundance of product reviews that businesses need to analyze, which can be time-consuming and resource-intensive. Additionally, the use of informal language, sarcasm, and irony in product reviews can make it challenging to accurately classify sentiments.

"Sentiment Analysis of Product Reviews Using Machine Learning Techniques" by R. K. Garg, et al. (2018). The paper presents a study on the application of machine learning techniques to sentiment analysis of product reviews. The authors collected data from various e-commerce websites and used several machine learning algorithms to classify reviews as positive, negative, or neutral. The study found that a combination of feature extraction techniques and machine learning algorithms can achieve high accuracy in sentiment analysis of product reviews.

"Real-Time Sentiment Analysis of Product Reviews on Social Media" by N. C. Mar, et al. (2017)
The paper presents a study on real-time sentiment analysis of product reviews on social media. The authors used a dataset of product reviews from Twitter and Facebook and used a combination of rule-based and machine learning techniques to classify reviews as positive, negative, or neutral. The study found that real-

time sentiment analysis of product reviews on social media can provide businesses with valuable insights into customer opinions and improve their customer service.

"Sentiment Analysis of Online Product Reviews: A Review" by M. K. Manchanda and M. K. Sharma (2019). The paper provides a comprehensive review of sentiment analysis of online product reviews. The authors discuss the importance of sentiment analysis in e-commerce and provide an overview of the various techniques used for sentiment analysis of product reviews. The study also highlights the challenges and limitations of sentiment analysis and suggests future research directions in the field.

"A Comparison of Supervised Machine Learning Algorithms for Sentiment Analysis of Product Reviews" by R. H. Gondal and S. F. Islam (2020). The paper presents a comparative study of supervised machine learning algorithms for sentiment analysis of product reviews. The authors used a dataset of product reviews from Amazon and compared the performance of several machine learning algorithms, including logistic regression, decision tree, and random forest. The study found that logistic regression outperformed other algorithms in terms of accuracy and efficiency.

"Real-Time Sentiment Analysis of Product Reviews Using Deep Learning" by P. S. Kim and Y. S. Kim (2019). The paper presents a study on real-time sentiment analysis of product reviews using deep learning techniques. The authors used a dataset of product reviews from Amazon and trained a deep neural network to classify reviews as positive, negative, or neutral. The study found that deep learning techniques can achieve high accuracy in sentiment analysis of product reviews and provide valuable insights into customer opinions.

Sentiment analysis of real-time product reviews is an important tool for businesses to understand their customers' opinions and improve their products and services. The studies reviewed in this literature review highlight the various techniques and challenges in sentiment analysis of product reviews. While machine learning techniques and deep learning techniques have shown promising results, there is still room for improvement in the accuracy and efficiency of sentiment analysis algorithms. Future research in the field should focus on developing more advanced techniques for sentiment analysis and addressing the challenges of analyzing real-time product reviews.

Chapter 3

Analysis and Design

3.1 Software Requirements

a. Visual Studio Code

Visual Studio Code is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

b. React

React is a JavaScript library to create user interfaces. Design simple views for each state in the application, and React will efficiently update and render just the right components when the data changes.

c. Redux

Redux is state container for JavaScript apps which is used to in-app state management in React. It helps us to centralize the application's state and logic enables powerful capabilities like undo/redo, state persistence.

d. CSS

CSS stands for Cascading Style Sheets and it is used to describe the presentation of a document written within a language like HTML.

e. JavaScript

JavaScript is a high-level, just-in-time compiled programming language that is one of the core technologies of the World Wide Web (WWW).

f. MUI

Material UI provides a robust, customizable, and accessible library of foundational and advanced components, enabling you to build the own design system and develop react applications faster.

g. Nodejs

Node is a asynchronous event-driven JavaScript runtime which is used to build scalable network applications.

h. Expressjs

Express is a backend web application framework for Node js. It is designed for building web applications and APIs.

3.2 Data Set

The training data set was taken from Kaggle. Dataset is a subset of Amazon Review 2018 dataset.Data used in this project includes consumer reviews for category Electronics. The data set contains19809 rows and 5 columns. The data set has the following attributes:

- overall - rating of the product (1 to 5)
- vote - helpful votes of the review
- reviewText - text of the review
- summary - summary of the review
- reviewTime - time of the review (raw)

Out of the 5 features, reviewText and overall were selected on which the model was trained.

```
df=pd.read_csv('electronics_sample.csv')
df.head()
```

	overall	vote	reviewTime	reviewText	summary
0	2	0	2010-02-10	Tech support is the worst	1265760000
1	2	0	2016-10-24	Screws were missing from the bracket and beaut...	Spend a little more and get much better.
2	1	0	2017-07-10	Trouble connecting and staying connected via b...	1499644800
3	4	5	2013-05-02	I purchased this unit for our RV to replace an...	Receiver Offers a Lot of Flexibility & Complexity
4	3	0	2013-01-04	It works. Nuff said but the review requires 1...	It's a cable

FIGURE 1: DATASET FROM KAGGLE

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19809 entries, 0 to 19808
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   overall     19809 non-null  int64
 1   vote        19809 non-null  int64
 2   reviewTime  19809 non-null  object
 3   reviewText  19808 non-null  object
 4   summary     19809 non-null  object
dtypes: int64(2), object(3)
memory usage: 773.9+ KB
```

FIGURE 2: ATTRIBUTES OF THE DATASET

3.3 Architecture

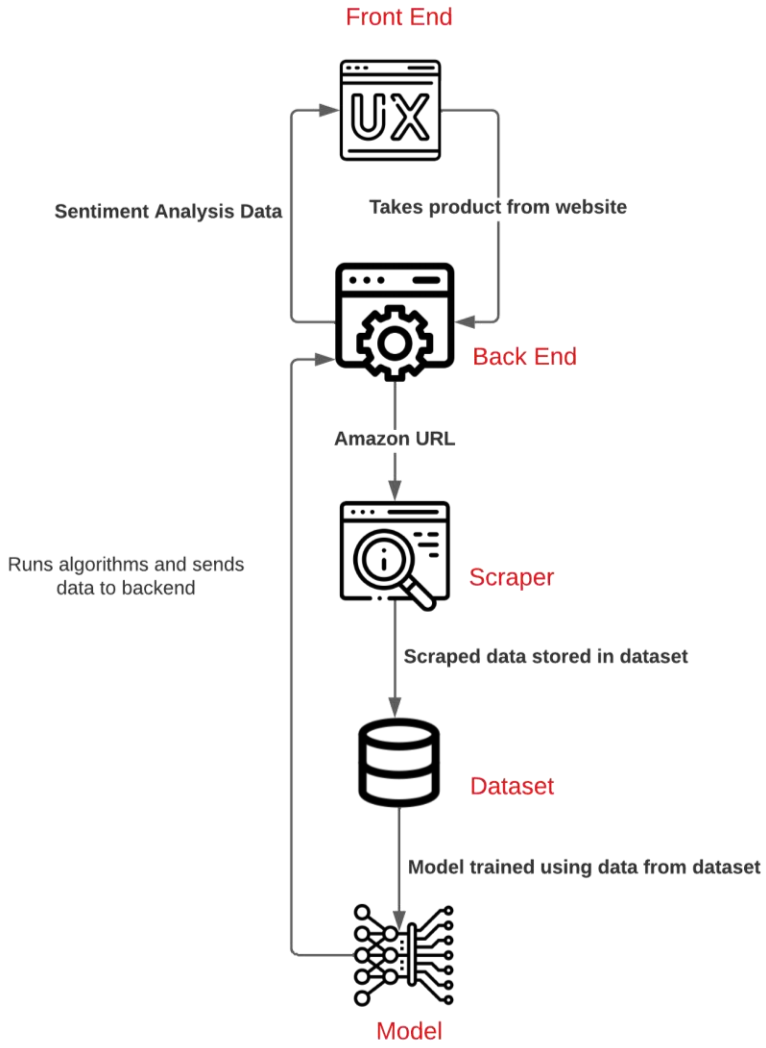

FIGURE 3: ARCHITECTURE

We developed a web application which predicts the sentiment of a product found on the amazon. This application uses three machine learning models to predict the sentiment of a review. The model is initially trained using an existing dataset of already identified sentiment with their reviews.

The user is initially provided with a dropdown where he can select a product. The selected product's url is sent to the backend which calls the scraper. Then, the scraper scrapes the given url and writes the data into a csv file. The three machine learning algorithms are run on the csv file which predict the sentiment of each review. Then, the csv file is converted in the form of json and sent back to the front-end where the results are displayed.

3.4 UML Diagram

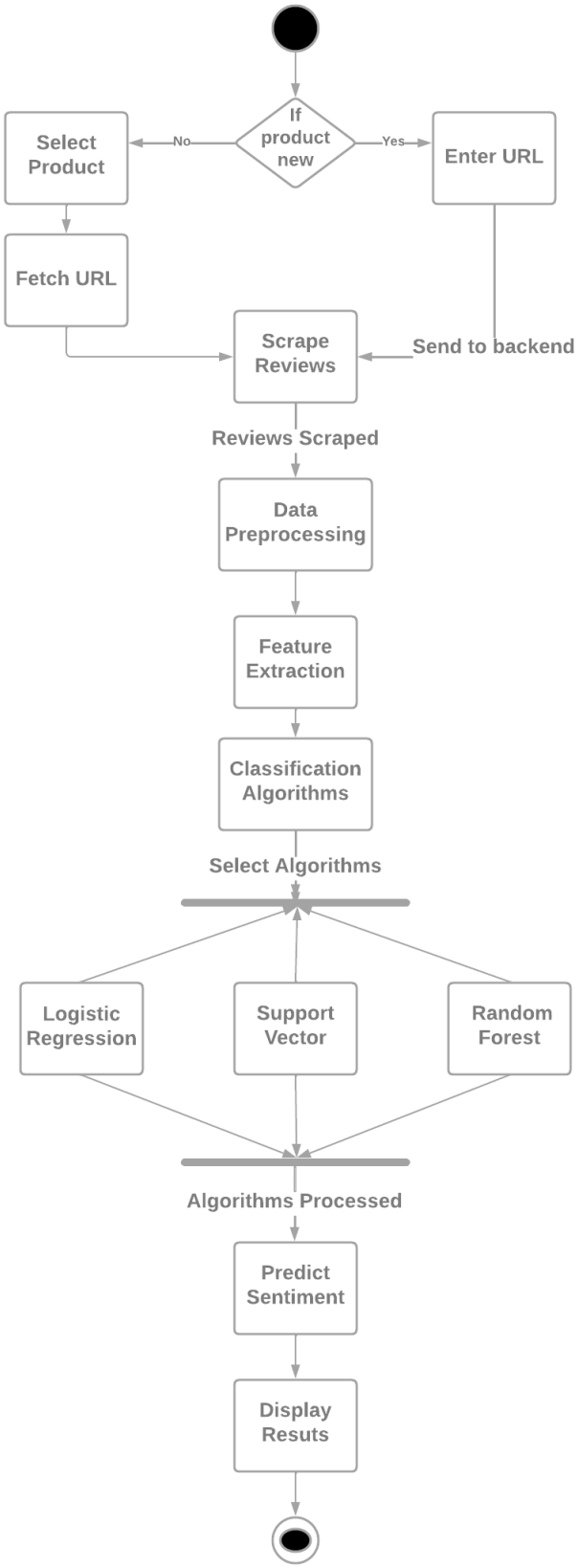

FIGURE 4: ACTIVITY DIAGRAM

1. User can run the sentiment analysis on a product from the given choices or he/she can feed in the amazon URL of a new product.

2. Once the product is selected then the reviews are scraped from amazon.

3. This scraped data is preprocessed. Preprocessing involves handling missing data, removing punctuations, removing stop words, tokenization and lemmatization.

4. Next the features are extracted using TF-IDF vectorizer.

5. Then 3 classification algorithms namely Support Vector Machine, Logistic Regression and Random Forest are applied on extracted features.

6. These algorithms predict the sentiment, 1 for positive and 0 for negative and send this data to the frontend.

7. Visualizations including overall sentiment and sentiment over time are displayed to the user.

3.5 Data Preprocessing

Pre-processing the data is the process of cleaning and preparing the text for classification. Onlinetexts usually contain lots of noise and uninformative parts. In addition, many words in the text donot have an impact on the general orientation of it. Pre-processing is done to reduce the noise in the text and it helps improve the performance of the classifier and speeds up the classification process. The whole process involves several steps:

3.5.1 Removing punctuations

The string.punctuation in python contains the following punctuation symbols.

!"#$%&\'()*+,-./:;<=>?@[\\]^_{|}~`. We can add or remove more punctuations as per our need.

	reviewText	text_wo_punct
7586	Only allows HD for one TV the other TV can onl...	Only allows HD for one TV the other TV can onl...
2860	I tried to install 4 Western Digital RED 3.0TB...	I tried to install 4 Western Digital RED 30TB ...
10369	Profit from my Unfortunate Experience with thi...	Profit from my Unfortunate Experience with thi...
10433	Canned Air - I Can't Recommend	Canned Air I Cant Recommend
11696	Great for creative gelling of studio lights. G...	Great for creative gelling of studio lights Go...

FIGURE 5: OUTPUT AFTER REMOVAL OF PUNCTUATION

3.5.2 Removing Stop words

After tokenizing, we remove stop words. Stop words are words used frequently but don't contributemuch to the meaning of the sentence, i.e., they are sentiment neutral.

	text_wo_punct	text_wo_stop
3196	Poor touch screen accuracy	poor touch screen accuracy
16324	it works	works
9206	Decent Surge Protector Poor USB Charger	decent surge protector poor usb charger
10843	Nice Drive Software Needs Work	nice drive software needs work
14194	You must need to be a mechanical engineer to f...	must need mechanical engineer figure secret co...

FIGURE 6: OUTPUT AFTER REMOVAL OF STOP WORDS

3.5.3 Stemming

The process of reducing inflected words to their word stem or root word. For every variation withevery root word more memory is required. With Stemming, we reduce the corpus of words the model is exposed to. A few of the Stemmers included in the nltk package are- Porter Stemmer, Snowball Stemmer, Lancaster Stemmer and Regex based Stemmer.

	text_wo_stop	text_stemmed
16029	im looking better one	im look better one
4234	tiny buttons tiny symbols instead words settin...	tini button tini symbol instead word set time ...
14001	sony scdce595 one great audiophile bargains ti...	soni scdce595 one great audiophil bargain time...
8770	entry level electronic device	entri level electron devic
12284	works great im glad	work great im glad

FIGURE 7: OUTPUT AFTER STEMMING

3.5.4 Lemmatization

Process of grouping together the inflected forms of a word so they can be analyzed as a single term, identified by the word's lemma, is called lemmatization. Lemmatization will always return a dictionary word and is typically more accurate, but is more computationally expensive.

	text_wo_stop	text_lemmatized
10027	great movie still camera	great movie still camera
11744	wasted money product still working fine yet we...	waste money product still work fine yet well d...
11412	modem 2 years already dead worked pretty well ...	modem 2 year already dead work pretty well til...
7400	amazed monitor bright clear use day long compu...	amaze monitor bright clear use day long comput...
18712	performs marginally better omnidirectional ant...	performs marginally well omnidirectional anten...

FIGURE 8: OUTPUT AFTER LEMMATIZATION

3.5.5 Vectorization

The process of encoding text as integers to create feature vectors is called vectorization. This process is done to convert text into a form that the machine learning model can understand. Different methods of vectorization are- Count vectorization, N-grams vectorization and Term frequency-Inverse document frequency vectorization (TF-IDF).

```
print(text)
```

```
  (0, 3405)     0.4888249860043193
  (0, 19607)    0.5593463763691049
  (0, 19948)    0.6694637886403032
  (1, 22155)    0.1628748207187583
  (1, 13513)    0.1974944030100606
  (1, 9189)     0.15288366500766853
  (1, 12214)    0.2050742913139871
  (1, 18962)    0.2787397129266404
  (1, 22135)    0.24465840754845067
  (1, 22402)    0.28029628968854886
  (1, 6783)     0.41954259338932776
  (1, 3609)     0.42687678488704095
```

FIGURE 9: OUTPUT AFTER VECTORIZATION

3.6 Algorithms

Three machine learning models were trained and used for sentiment classification:

3.6.1 Logistic Regression

Logistic regression is one of the supervised machine learning algorithms. It is used for predictingthe categorical dependent variable using a given set of independent variables. Logistic regressionpredicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. Logistic Regression is much similar to Linear Regression except thathow they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems. In Logistic regression, instead of fittinga regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or1). The curve from the logistic function indicates the likelihood of something. Logistic Regressionis a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets. The sigmoid function is a mathematical function used to map the predicted values to probabilities. It maps any real value into another valuewithin a range of 0 and 1. The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.

```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(solver='liblinear', random_state=1)
classifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
y_pred_tr = classifier.predict(x_train)
from sklearn.metrics import classification_report
print("Classification Report(Train)")
print(classification_report(y_train, y_pred_tr))
print("Classification Report(Test)")
print(classification_report(y_test, y_pred))
```

```
Classification Report(Train)
              precision    recall  f1-score   support

           0       0.85      0.92      0.88      9460
           1       0.87      0.76      0.81      6386

    accuracy                           0.85     15846
   macro avg       0.86      0.84      0.85     15846
weighted avg       0.86      0.85      0.85     15846

Classification Report(Test)
              precision    recall  f1-score   support

           0       0.80      0.89      0.84      2360
           1       0.81      0.67      0.73      1602

    accuracy                           0.80      3962
   macro avg       0.80      0.78      0.79      3962
weighted avg       0.80      0.80      0.80      3962
```

FIGURE 10: CLASSIFICATION REPORT OF LOGISTIC REGRESSION

3.6.2 Support Vector Machine

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine. There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

```
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
svm = LinearSVC(C=0.075)
svm.fit(x_train,y_train)
y_pred_tr = svm.predict(x_train)
print("Classification Report(Train)")
print(classification_report(y_train, y_pred_tr))
print("Classification Report(Test)")
print(classification_report(y_test, y_pred))
```

```
Classification Report(Train)
              precision    recall  f1-score   support

           0       0.83      0.92      0.88      9460
           1       0.86      0.73      0.79      6386

    accuracy                           0.84     15846
   macro avg       0.85      0.82      0.83     15846
weighted avg       0.85      0.84      0.84     15846

Classification Report(Test)
              precision    recall  f1-score   support

           0       0.80      0.89      0.84      2360
           1       0.81      0.67      0.73      1602

    accuracy                           0.80      3962
   macro avg       0.80      0.78      0.79      3962
weighted avg       0.80      0.80      0.80      3962
```

FIGURE 11: CLASSIFICATION REPORT OF SVM

3.6.3 Random Forest

It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Random Forest is a classifier that contains a number of decision trees on various subsets of the given datasetand takes the average to improve the predictive accuracy of that dataset. Instead of relying on onedecision tree, the random forest takes the prediction from each tree and based on the majority votesof predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

```
from sklearn.ensemble import RandomForestClassifier
classifier = RandomForestClassifier(min_samples_leaf=5)
classifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
y_pred_tr = classifier.predict(x_train)
from sklearn.metrics import classification_report
print("Classification Report(Train)")
print(classification_report(y_train, y_pred_tr))
print("Classification Report(Test)")
print(classification_report(y_test, y_pred))
```

```
Classification Report(Train)
              precision    recall  f1-score   support

           0       0.81      0.93      0.87      9460
           1       0.87      0.68      0.76      6386

    accuracy                           0.83     15846
   macro avg       0.84      0.81      0.82     15846
weighted avg       0.84      0.83      0.83     15846

Classification Report(Test)
              precision    recall  f1-score   support

           0       0.77      0.91      0.84      2360
           1       0.82      0.60      0.69      1602

    accuracy                           0.79      3962
   macro avg       0.80      0.75      0.76      3962
weighted avg       0.79      0.79      0.78      3962
```

FIGURE 12: CLASSIFICATION REPORT OF RANDOM FOREST

3.7 Web Scraping

Web scraping is a technique used to extract large amount of data from websites and store it in your computer. This data can be later used for analysis. The library requests is used to get the content from a web page. We send a request to the URL and we get a response. The response will contain a status code along with the web page content. BeautifulSoup converts the contents of a page into a proper format.

STEPS:

1. Get the URL of the page to be scrapped.
2. Inspect the elements of the page and identify the tags required.
3. Access the URL.
4. Get the element from the required tags.

Once we have scrapped all the reviews, we have to save it in a file in order to perform further analysis. We convert the reviews list into a dictionary. Then import the pandas library and use it to convert the dictionary into a data frame. Then using to_csv() function we convert it into a CSV file and do continue our analysis.

	name	date	title	review	vote	rating
0	Somesh Saraf	2019-10-30	Quite disappointed!!	Writing a review after few days of usage.It's ...	511	1
1	Hari Parkash	2019-10-18	Premium! BUT Not for Middle class	Very honest review, when I saw the price of th...	343	3
2	Pradip Kumar Paul	2019-10-20	Fantastic and all rounder band.	Fantastic and all rounder band. It's all in one.	104	5
3	Lokesh Pawar	2020-08-20	Un satisfied	I am Mr Aniket Pawar! had purchased the smart ...	48	1
4	Mat	2019-11-11	Worth or not worth based on individuals. I wou...	I liked it before I saw the Samsung gear which...	57	2
5	Abhishek Tiwari	2020-09-23	Not as expected 😠	Looks good👍I think I expected more from FitBi...	21	3
6	ashish kumar	2020-02-18	Writing the review after 1 month of use. See f...	Pros:1. The watch looks attractive with a good...	20	4
7	Anil T.	2019-10-31	Software issues	Some software issues with the app. It get conn...	21	1
8	Amazon Customer	2020-10-21	Please read my reveiw and then decide.	I bought such an expensive watch mainly to tra...	13	1
9	hitesh bhalodia	2019-11-04	Very good product go for it..	It is very good product. I used in swimming p...	17	4

FIGURE 13: OUTPUT AFTER SCRAPING

3.8 Front End

The predicted data is then converted into json which is sent back to the front end. Here, React is used to display all the results such as pie charts, bar charts and line charts. Sentiment highlights of a few random reviews are also shown.

Chapter 4

Implementation and Result Discussion

4.1 Screenshots

This is the landing page where the user can see the sentiment summary of all the products. The user can select any product individually to get the sentiment analysis of that product. The user can do analysis of a new product by clicking on Add Keyword. Next the user will be asked to enter the amazon url of the product. This type of summary is useful for companies that have a range of product.

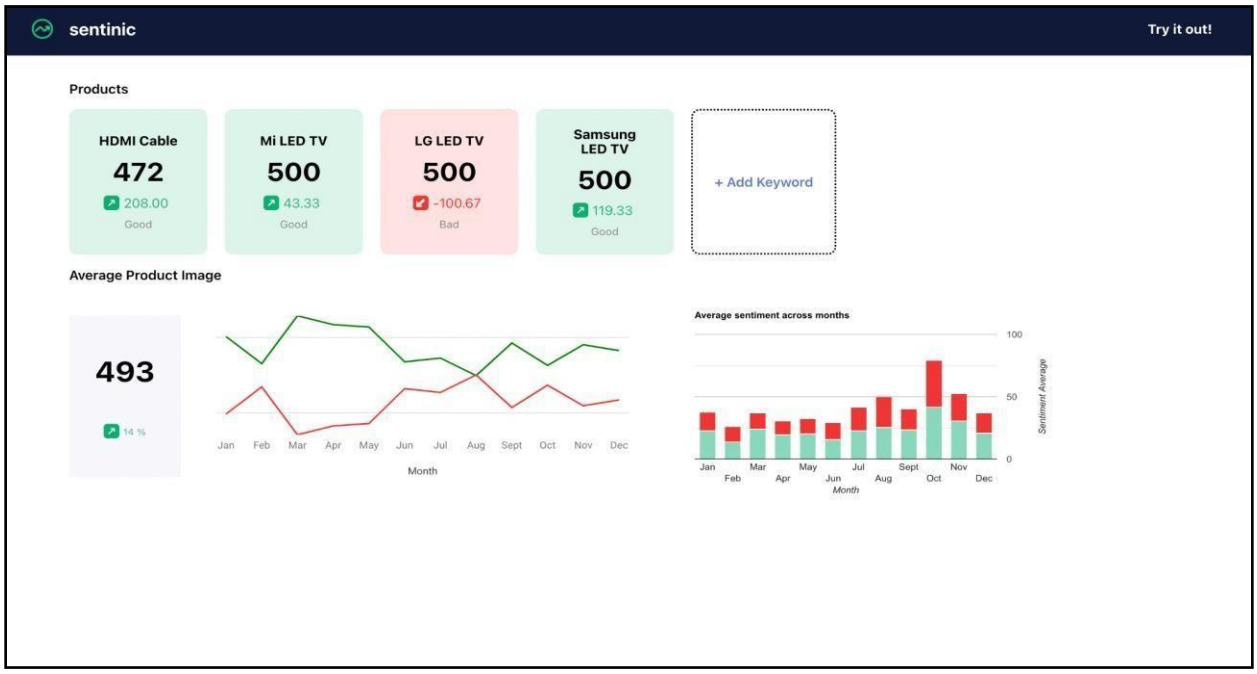

FIGURE 14: AVERAGE SENTIMENT OF ALL PRODUCTS

Once we have selected a particular product, then we can view the in-depth analysis of that product and its performance using the three machine learning models that we have implemented. If we select the first classifier that is Logistic Regression, then we can see the sentiment analysis of the product over the months through the graph and the bar chart. The difference between the positive and negative sentiments is 28% as predicted by Logistic Regression. The first ten product reviews are also highlighted on the right side with their respective sentiments.

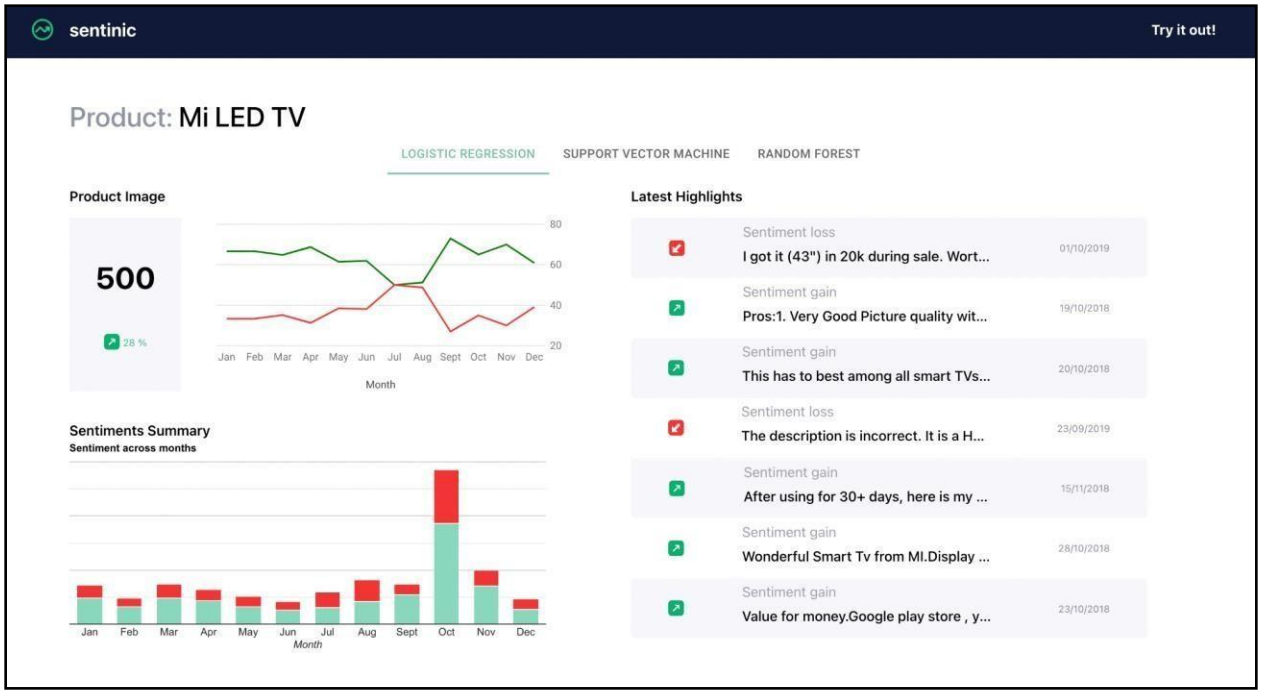

FIGURE 15: SENTIMENT DETAILS USING LOGISTIC REGRESSION

This graph depicts the percentage of positive and negative reviews. So the product in reference has 60% positive reviews and 40% negative reviews.

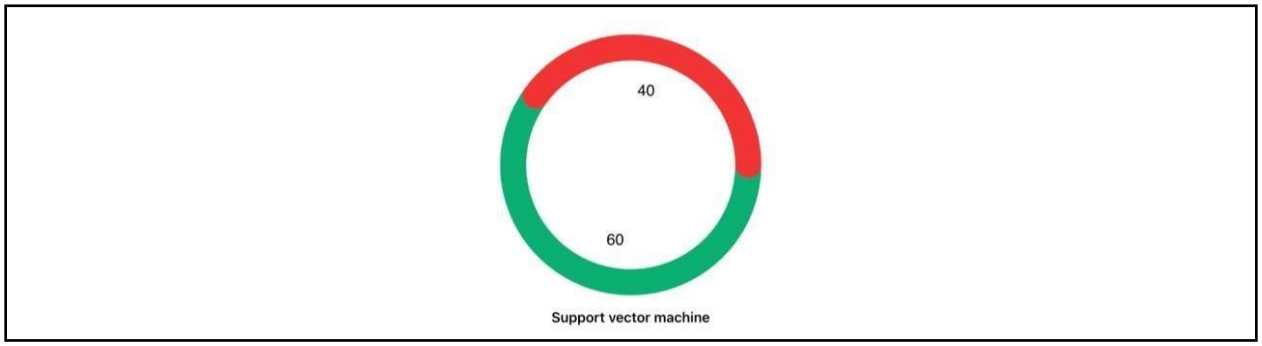

FIGURE 16: SENTIMENT PIE-CHART

For our second classifier that is Support Vector Machine, we get a similar analysis and we see that the difference between the negative and positive sentiments is around 21%.

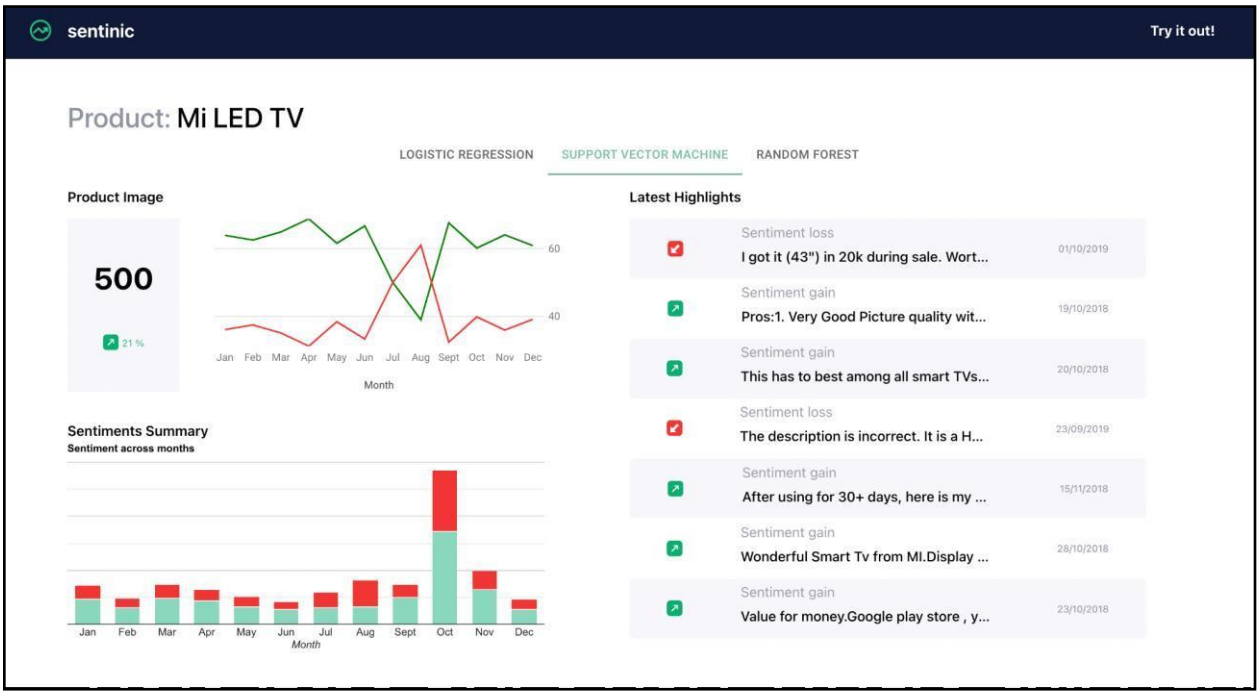

FIGURE 17: SENTIMENT DETAILS USING SVM

In Random Forest, we can see that the negative sentiments are 23% more than positive sentiment.

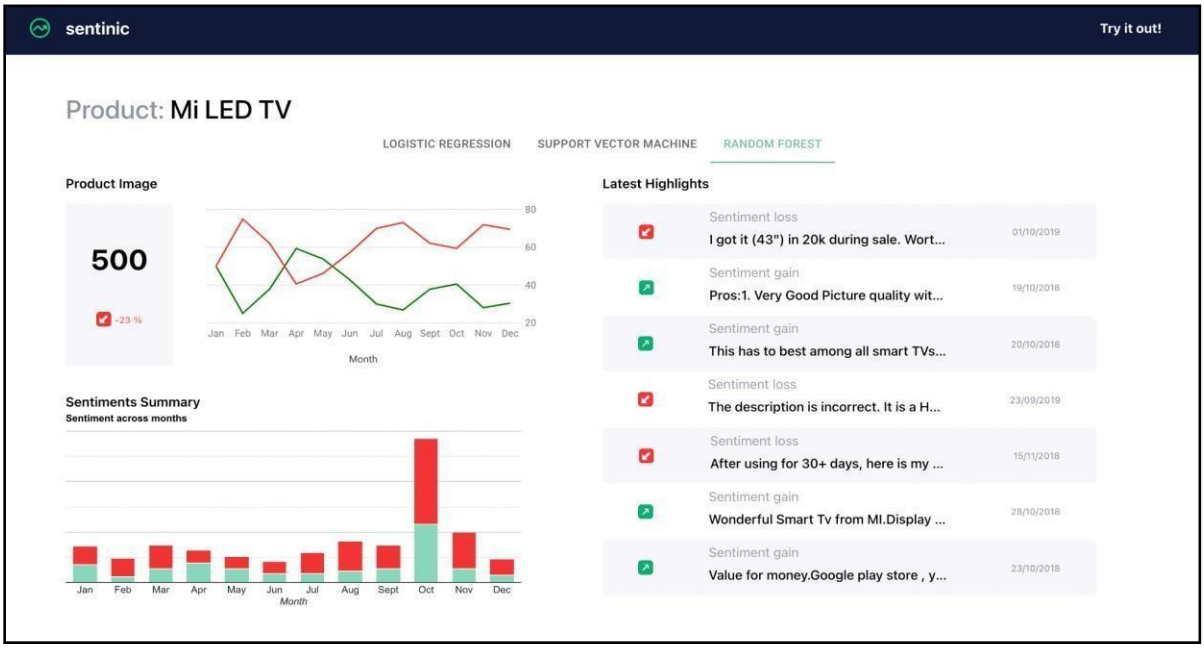

FIGURE 18: SENTIMENT DETAILS USING RANDOM FOREST

Chapter 5

Conclusion and Future Work

Businesses play an important role in nation building. As ecommerce is gaining popularity, it becomes very important for the business owners to better understand their customers. Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helps a business to understand the social sentiment of their brand, product or service while monitoring online reviews. The reviews provided by the customers on e-commerce websites not only helps the owners to improve their services accordingly but also helps other customers to get an opinion of the products before buying them online. Our project was aimed at developing a web application to perform sentiment analysis on amazon product reviews by leveraging machine learning algorithms.

In order to analyze the customer reviews and the sentiment attached to the reviews, in the past sentiment analysis has been performed on various datasets. We reviewed 8 such papers that compared different classification algorithms like Random Forest, Support Vector Machine, Naïve Bayes, Decision Tree and Logistic Regression. We analyzed the features and drawbacks of these papers and selected 3 algorithms that performed best – Logistic regression, Random forest and SVM.

We developed a web application which predicts the sentiment of a product found on amazon. This application uses three machine learning models to predict the sentiment of a review. The model is initially trained using an existing dataset of already identified sentiment with their reviews. The user is initially provided with a dropdown where he can select a product. The selected product's url is sent to the backend which calls the scraper. Then, the scraper scrapes the given url and writes the data into a csv file. The three machine learning algorithms are run on the csv file which predicts the sentiment of each review. Then the csv file is converted in the form of json and sent back to the front-end where the results are displayed.

Businesses can utilize our sentiment analysis web application to work on their product's incoming reviews 24/7 and gain instant insights from their customer data. Using our striking visualizations, they can easily track the performance of their product range in the market. Product teams often get caught up in day-to-day tasks and forget to listen to what the customer is saying. Sometimes, they don't even receive product feedback because companies don't have a feedback loop system in

place. However, once that data starts flowing in, it is important to make sure that the data is analyzed in a fast, accurate, and cost-effective way.

In future, we plan to make our web application more sophisticated by using aspect based sentiment analysis and fake review detection. We also want to go a bit further to build a product recommendation system and generalize the application for any reviews like movie or food reviews. We will also take into consideration certain quality measures like review helpfulness to improve the model accuracy.

References

[1] Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019).Sentiment analysis on product reviews using machine learning techniques. In Cognitive Informatics and Soft Computing (pp. 639-647). Springer, Singapore.

[2] Safrin, R., Sharmila, K. R., Subangi, T. S., & Vimal, E. A. (2017). Sentiment analysis on online product review.Int. Res. J. Eng. Technol, 4(04).

[3] Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1), 1-14.

[4] Sultana, N., Kumar, P., Patra, M. R., Chandra, S., & Alam, S. (2019). Sentiment analysis for product review.Int. J. Soft Comput, 9(7).

[5] Singla, Z., Randhawa, S., & Jain, S. (2017, June). Sentiment analysis of customer product reviews using machine learning. In 2017 international conference on intelligent computing and control (I2C2) (pp. 1-5). IEEE.

[6] Ghosh, S., Hazra, A., & Raj, A. (2020). A Comparative Study of Different Classification Techniques for Sentiment Analysis. International Journal of Synthetic Emotions (IJSE), 11(1), 49-57.

[7] Singh, S. N., & Sarraf, T. (2020, January). Sentiment analysis of a product based on user reviews using random forests algorithm. In 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 112-116). IEEE.

[8] Ramdhani, S. L., Andreswari, R., & Hasibuan, M. A. (2018, November). Sentiment analysis of product reviews using naive bayes algorithm: A case study. In 2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT) (pp. 123- 127). IEEE.

[9] Zvarevashe, K., & Olugbara, O. O. (2018, March). A framework for sentiment analysis with opinion mining of hotel reviews. In 2018 Conference on information communications technology and society (ICTAS) (pp. 1- 4). IEEE.

[10] Nair, A. J., Veena, G., & Vinayak, A. (2021, April). Comparative study of Twitter Sentiment On COVID-19 Tweets. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1773-1778). IEEE.

[11] Haque, M. R., Lima, S. A., & Mishu, S. Z. (2019,December). Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews. In 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) (pp. 161-164). IEEE.

[12] Yadav, S., & Saleena, N. (2020, October). Sentiment Analysis Of Reviews Using an Augmented Dictionary Approach. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp.1-5). IEEE.

[13] Haberzettl, M., & Markscheffel, B. (2018). A Literature Analysis for the Identification of Machine Learning and Feature Extraction Methods for Sentiment Analysis. In ICDIM (pp. 6-11).

[14] Han, K. X., Chiu, C. C., & Chien, W. (2019, October). The Application of Support Vector Machine (SVM) on the Sentiment Analysis of Internet Posts. In 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE) (pp. 154-155). IEEE.

[15] Smetanin, S., & Komarov, M. (2019, July). Sentiment analysis of product reviews in Russian

usingconvolutional neural networks. In 2019 IEEE 21st Conference on Business Informatics (CBI) (Vol. 1, pp. 482-486). IEEE.

[16] Bandana, R. (2018, May). Sentiment analysis of movie reviews using heterogeneous features. In 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) (pp. 1-4). IEEE.

[17] Nafees, M., Dar, H., Lali, I. U., & Tiwana, S. (2018, November). Sentiment analysis of polarity in product reviews in social media. In 2018 14th International Conference on Emerging Technologies (ICET) (pp. 1-6). IEEE.

[18] Kumari, U., Sharma, A. K., & Soni, D. (2017, August). Sentiment analysis of smart phone product review using SVM classification technique. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 1469-1474). IEEE.

[19] Hidayah, I., Permanasari, A. E., & Wijayanti, N. W. (2019, July). Sentiment Analysis on Product Review using Support Vector Machine (SVM). In 2019 5th International Conference on Science and Technology (ICST) (Vol. 1, pp. 1-4). IEEE.

[20] Shah, B. K., Jaiswal, A. K., Shroff, A., Dixit, A. K., Kushwaha, O. N., & Shah, N. K. (2021, January). Sentiments Detection for Amazon Product Review. In 2021 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.

[21] Noor, A., & Islam, M. (2019, July). Sentiment Analysis for Women's E-commerce Reviews using Machine Learning Algorithms. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

[22] Alrehili, A., & Albalawi, K. (2019, April). Sentiment analysis of customer reviews using ensemble method. In 2019 International Conference on Computer and Information Sciences (ICCIS) (pp. 1- 6). IEEE.

[23] Salem, M. A., & Maghari, A. Y. (2020, December). Sentiment Analysis of Mobile Phone Products Reviews Using Classification Algorithms. In 2020 International Conference on Promising Electronic Technologies (ICPET) (pp. 84-88). IEEE.

[24] Faisol, H., Djajadinata, K., & Muljono, M. (2020, September). Sentiment Analysis of Yelp Review. In 2020 International Seminar on Application for Technology of Information and Communication (iSemantic) (pp. 179-184). IEEE.

[25] Kaur, H., & Mangat, V. (2017, February). A survey of sentiment analysis techniques. In 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 921-925). IEEE.

[26] Sudhir, P., & Suresh, V. D. (2021). Comparative study of various approaches, applications and classifiers for sentiment analysis. Global Transitions Proceedings.

[27] S. Gupta, Sentiment Analysis: Concept, Analysis and Applications, Towards Data Science, Jan. 7 2018. Accessed on: Dec. 2, 2021. [Online]. Available: https://towardsdatascience.com/sentiment-analysisconcept-analysis-and-applications- 6c94d6f58c17

[28] Ni, J., Li, J., & McAuley, J. (2019, November). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing

Acknowledgment

We would like to thank our college SVKM ‘s NMIMS Mukesh Patel School of Technology Management and Engineering for introducing this Technical Project which exposed us to the practical implementation of Computer Science Technology besides gave us a platform to have hands-on experience to learn new technologies and apply them. We submit my thanks and respect to the Department of Computer Science, NMIMS for driving us to explore such a wonderful opportunity.

My faculty mentor Dr. Pravin Shrinath has given us constant support and guidance besides being a source of inspiration through his words of encouragement. His continuous supervision and much-needed guidance have really helped us in conceptualizing our project and executing the same. We express our gratitude to Dr. Alka Mahajan (Dean MPSTME - Mumbai) for her encouragement and for providing a stimulating environment of education at our institute.

The project has given several valuable inputs on the design, development, and execution of the project besides introducing us to contemporary and new technologies. We are extremely grateful to all the technical and HR staff of NMIMS MPSTME for their cooperation and guidance that has helped us a lot during training. We have learned a lot working under them and we will always be thankful to them for this valuable addition to me.

We express my gratitude to my parents for their support, blessings, and best wishes.

Romit Changani

Pratyaksh Jain

Tanish Korgaonkar

Karthik Ram Srinivas

B. Tech - Computer Science and Business Systems

Mukesh Patel School of Technology Management & Engineering NMIMS, Mumbai