AN INTERNSHIP REPORT ON

Amazon Alexa Reviews Sentimental Analysis A Report Submitted to

Blackbuck Engineers Pvt. Ltd

Submitted by
T. LAKSHMI NAVEENA (BBRCP594)

of
Department of CSE (CYBER SECURITY)



Blackbuck Engineers Pvt. Ltd Road no 36, Jubilee Hills, Hyderabad





Internship Experience Letter

Certificate ID: BBNR0054. Issued Date: 21stJune, 2023

To Whom It May Concern:

This is to certify that T. LAKSHMI NAVEENA (BBRCP594) has successfully completed the internship at BLACKBUCK ENGINEERS Pvt Ltd Hyderabad from 15th May 2023 to 17th June 2023.

He has worked on a project titled **AI Voice Assistant** by learning and incorporating Artificial Intelligence & Machine Learning concepts under the supervision of our project mentor.

We found that he is sincere, hardworking, technically sound and result oriented. He worked well as part of a team during his tenure.

We wish him all the best for his future endeavors

Best regards,

Kathyayani. R

Project Head

Houris and Engage

Mounika Bezawada

HR Manager







ACKNOWLEDGEMENT

I wish to take this opportunity to express my deep gratitude to all the people who have extended their cooperation in various ways during my project work. It is my pleasure and responsibility to acknowledge the help of all those individuals.

I sincerely thank our guide **Ms. Vindya vasini**, for helping me in successful completion of our project under her supervision.

I thank the project coordinators **R. Kathyayani** and **B. Raj Kumar**, for their valuable guidance and support throughout the development of this project.

I express my deepest gratitude to **Anuradha Thota**, Founder and CEO,Blackbuck Engineers for her support and encouragement.

Finally, I thank one and all who directly or indirectly helped me to complete my project work successfully.

T. LAKSHMI NAVEENA (20ME1A4651)

ABSTRACT

Amazon Alexa is a popular voice-activated intelligent personal assistant that has gained significant popularity in recent years. As a result, there are a large number of reviews on Amazon.com related to Amazon Alexa that provide valuable insights into users' opinions and sentiments about the product. Sentiment analysis can be applied to these reviews to extract useful insights and understand how users feel about the product.

The objective of performing sentiment analysis on Amazon Alexa reviews is to identify the overall sentiment of the reviews as positive, negative, or neutral. This can be achieved by analyzing the language used in the reviews and determining the sentiment expressed by the words and phrases used. Sentiment analysis can also provide insights into the specific features and aspects of the Amazon Alexa product that users like or dislike

Sentiment analysis of Amazon Alexa reviews can be performed using various machine learning techniques, such as TextBlob or VaderSentiment libraries. These techniques involve preprocessing textual data, converting the text to a numerical representation, and training machine learning algorithms to predict the sentiment of new reviews.

Overall, sentiment analysis of Amazon Alexa reviews provides valuable insights into user sentiments and preferences regarding the product. This information can be used by Amazon to improve the product and provide a better user experience.

INDEX

CONTENTS

1. Introduction1	
2. Analysis4	
3. Software requirements specifications	
4. Technologies1	0
Python	
Collab	
Numpy	
Pandas	
Support vector machine	
5. Coding	. 13
6. Screenshots	.17
7. Conclusion	21
8. Bibliography	23

Learning Objectives/Internship Objectives

- Internships are generally thought of to be reserved for college students looking to gain experience in a particular field. However, a wide array of people can benefit from TrainingInternships in order to receive real world experience and develop their skills.
- An objective for this position should emphasize the skills you already possess in the area and your interest in learning more
- Internships are utilized in a number of different career fields, including architecture, engineering, healthcare, economics, advertising and many more.
- Some internship is used to allow individuals to perform scientific research while others are specifically designed to allow people to gain first-hand experience working.
- ➤ Utilizing internships is a great way to build your resume and develop skills that can be emphasized in your resume for future jobs. When you are applying for a TrainingInternship, make sure to highlight any special skills or talents that can make you stand apartfromthe rest of the applicants so that you have an improved chance of landing the position.

WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

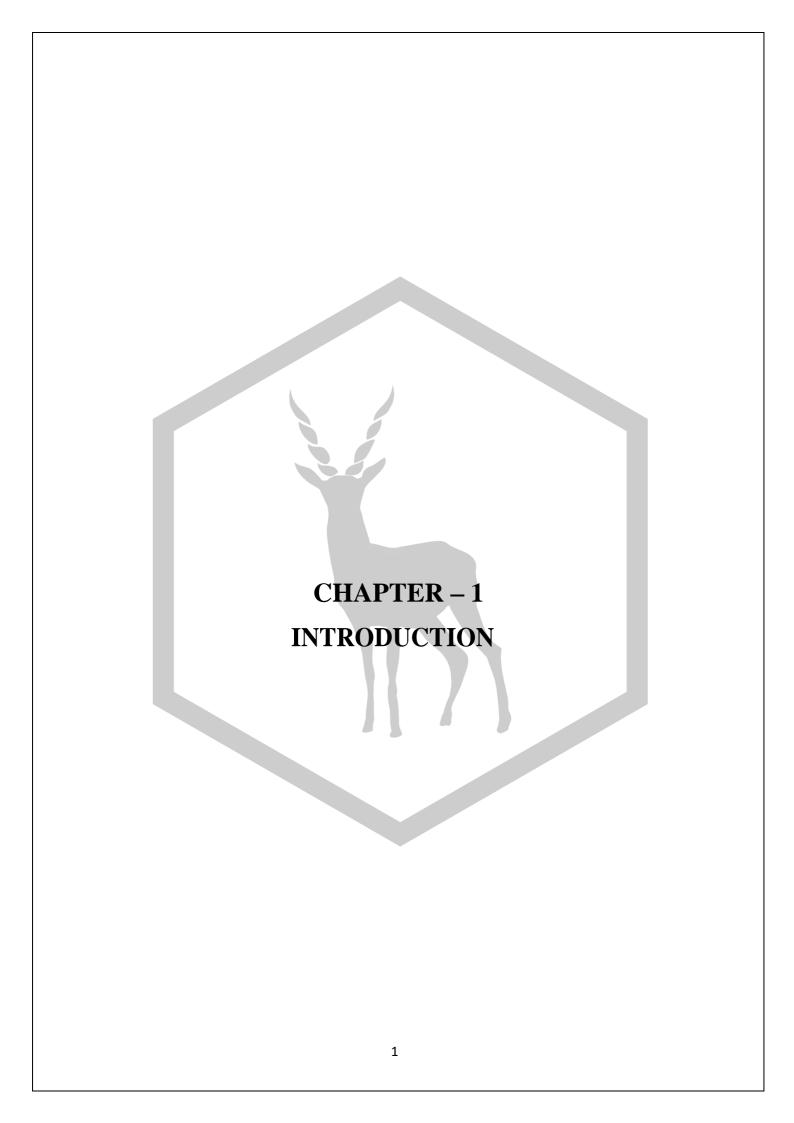
	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED	
	15.05.2023	Monday	Onboarding the Students & Introduce the Topic	
X	16.05.2023	Tuesday	Session-1 Basics on Python	
1st WEEK	17.05.2023	Wednesday	Session-2 MI Supervised and Unsupervised Algorithm	
1st '	18.05.2023	Thursday	Session-3 Deep Learning	
	19.05.2023	Friday	Session-4 Project Preprocessing	
	20.05.2023	Saturday	Session-5 Model Training	

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED		
	22.05.2023	Monday	Session-6 Model Project Development & Process		
M			Explanation		
VEEK	23.05.2023	Tuesday	Abstract Building		
WE	24.05.2023	Wednesday	Abstract Building		
2nd v	25.05.2023	Thursday	Abstract Building		
7	26.05.2023	Friday	Abstract Submission		
	27.05.2023	Saturday	Abstract Submission		

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED	
	29.05.2023	Monday	PPT Preparation	
\mathbf{X}	30.05.2023	Tuesday	PPT Preparation	
Œ	31.05.2023	Wednesday	PPT Preparation	
WEE				
3rd	01.06.2023	Thursday	PPT Preparation	
3				
	02.06.2023	Friday	PPT Preparation	
	03.06.2023	Saturday	Mid Review	
,			3.37	

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED	
	05.06.2023	Monday	Mid Review	
4thWEEK	06.06.2023	Tuesday	Mid Review	
WE	07.06.2023	Wednesday	Mid Review	
th	08.06.2023	Thursday	Coding and Documentation	
4	09.06.2023	Friday	Coding and Documentation	
	10.04.2023	Saturday	Coding and Documentation	

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
l ⊾a	12.06.2023	Monday	Coding and Documentation
EK	13.06.2023	Tuesday	Coding and Documentation
WE	14.06.2023	Wednesday	Final Review
5 th V			
S	15.06.2023	Thursday	Final Review
	16.06.2023	Friday	Final Review
	17.06.2023	Saturday	Final Review



INTRODUCTION

Amazon Alexa is a popular virtual assistant developed by Amazon that can help users with a variety of tasks such as setting reminders, playing music, making phone calls, and controlling smart home devices. As the use of Amazon Alexa continues to grow, there has been a growing interest in analyzing user reviews to gain insights into how people perceive and use the device. Sentiment analysis is a technique used to automatically determine the sentiment expressed in a piece of text, such as a user review. The goal of sentiment analysis is to classify the text as positive, negative, or neutral, based on the sentiment conveyed by the words and phrases used.

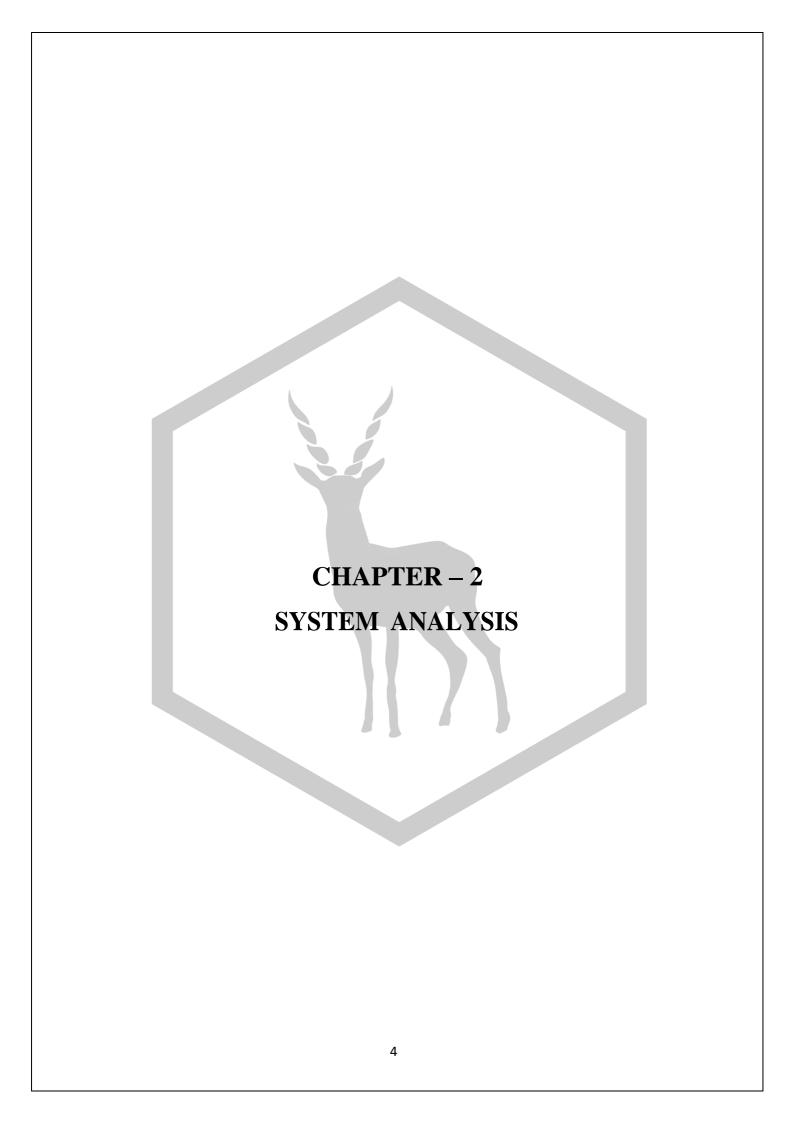
Some of the potential applications of Amazon Alexa reviews sentiment analysis include:

- Identifying common complaints or issues users have with the device
- Assessing the overall satisfaction level of users with the device
- Analyzing the sentiment towards specific features or capabilities of the device
- Comparing the sentiment of Alexa reviews to other virtual assistant reviews, such as Google Assistant or Siri.

Amazon Alexa reviews sentiment analysis involves analyzing the sentiments expressed in the user reviews of Amazon Alexa. The analysis can provide valuable insights into how people perceive the device and its features, as well as identify areas for improvement. With sentiment analysis, it is possible to identify common complaints or issues users have with the device, assess the overall satisfaction level of users, analyze the sentiment towards specific features or capabilities of the device, and compare the sentiment of Alexa reviews to other virtual assistant reviews, such as Google Assistant or Siri.

There are various techniques used for sentiment analysis, including supervised learning, unsupervised learning, and deep learning. Supervised learning involves training a machine learning algorithm using labeled data, where each review is labeled with its corresponding sentiment. Unsupervised learning techniques, such as clustering or topic modelling, can be used to identify patterns in the text that indicate positive, negative, or neutral sentiment. Deep learning techniques, such as neural networks, can also be used for sentiment analysis, which can be trained to identify patterns in the text that indicate sentiment.

The insights gained from Amazon Alexa reviews sentiment analysis can be used to improve the user experience, inform product development decisions, and identify areas for future research. For example, if the sentiment analysis identifies common complaints or issues users have with the device, these can be addressed in future updates or new product releases. Additionally, the sentiment analysis can be used to assess the overall satisfaction level of users with the device and identify areas for improvement. Analyzing the sentiment towards specific features or capabilities of the device can also provide valuable insights into which features are most valued by users and which may need further development. Overall, Amazon Alexa reviews sentiment analysis can provide valuable insights into how users perceive and use the device, which can be used to improve the user experience and inform future product development decisions.



SYSTEM ANALYSIS

EXISTING SYSTEM

- Amazon Comprehend: Amazon Comprehend is a fully managed NLP service provided by Amazon Web Services. It can be used to analyze Amazon Alexa reviews to determine sentiment, as well as to identify key phrases, entities, and topics in the text. Amazon Comprehend uses machine learning algorithms to automatically identify patterns in the text and classify the sentiment as positive, negative, or neutral. Users can also train custom models to classify sentiment based on their own criteria.
- MonkeyLearn: MonkeyLearn is a cloud-based machine learning platform that
 provides pre-built sentiment analysis models for analyzing Amazon Alexa reviews.
 The platform uses supervised learning algorithms to classify the sentiment expressed
 in the text, and provides an intuitive user interface for training and testing the models.
 MonkeyLearn also offers integrations with popular tools like Zapier and Google
 Sheets, allowing users to easily automate the sentiment analysis process.
- IBM Watson Natural Language Understanding: IBM Watson Natural Language Understanding is an NLP service that can be used to analyze the sentiment expressed in Amazon Alexa reviews. The service can detect positive, negative, and neutral sentiment, and can also identify entities, categories, concepts, and emotions in the text. IBM Watson Natural Language Understanding uses machine learning algorithms to analyze the text and provide insights into the sentiment expressed in the reviews.
- Lexalytics: Lexalytics is a text analytics platform that provides sentiment analysis capabilities for Amazon Alexa reviews. The platform uses a combination of rule-based and machine learning techniques to classify the sentiment expressed in the text. Lexalytics also provides visualization and reporting tools for analyzing the results, including dashboards, heat maps, and graphs.
- Google Cloud Natural Language: Google Cloud Natural Language is an NLP service
 that can be used to analyze Amazon Alexa reviews to determine sentiment. The
 service can detect positive, negative, and neutral sentiment, as well as identify
 entities, sentiment modifiers, and syntactic relationships in the text. Google Cloud
 Natural Language uses machine learning algorithms to analyze the text and provide

- insights into the sentiment expressed in the reviews.
- These existing systems and tools provide a range of options for performing sentiment analysis on Amazon Alexa reviews. Depending on the specific needs and requirements of the analysis, different tools and techniques may be more appropriate. Additionally, there are many other sentiment analysis tools and services available that can be used to analyze Amazon Alexa reviews, including open source libraries like NLTK and spaCy, as well as commercial products like RapidMiner and SAS.

PROPOSED SYSTEM

- Data Collection: The first step in any sentiment analysis project is to collect data. In this case, we would need to collect a large sample of Amazon Alexa reviews. This data can be collected either by scraping reviews from Amazon's website, or by using an API to access the reviews directly.
- Data Preprocessing: Once the data has been collected, it needs to be preprocessed to
 prepare it for analysis. This might involve tasks like removing punctuation,
 converting all text to lowercase, and removing stop words. The data may also need to
 be cleaned to remove irrelevant or spam reviews.
- Sentiment Analysis: The next step is to perform sentiment analysis on the preprocessed data. This can be done using one of the existing systems or tools described earlier, or by developing a custom sentiment analysis model. The model would need to be trained on a labeled dataset of Amazon Alexa reviews to accurately classify the sentiment as positive, negative, or neutral.
- Visualization: After performing sentiment analysis, the results can be visualized to
 provide insights into the sentiment expressed in the reviews. This might involve
 creating charts or graphs to show the distribution of positive, negative, and neutral
 sentiment, or to show the most common topics or keywords mentioned in the reviews.
- Reporting: Finally, the results of the sentiment analysis can be used to generate reports or dashboards to help stakeholders understand the sentiment expressed in the Amazon Alexa reviews. These reports might include information like overall

- sentiment trends, sentiment by product or feature, or sentiment by customer demographics.
- Overall, a proposed system for Amazon Alexa reviews sentiment analysis would involve collecting and preprocessing data, performing sentiment analysis, visualizing the results, and generating reports to help stakeholders make informed decisions based on customer sentiment. The specific tools and techniques used will depend on the needs and requirements of the project.





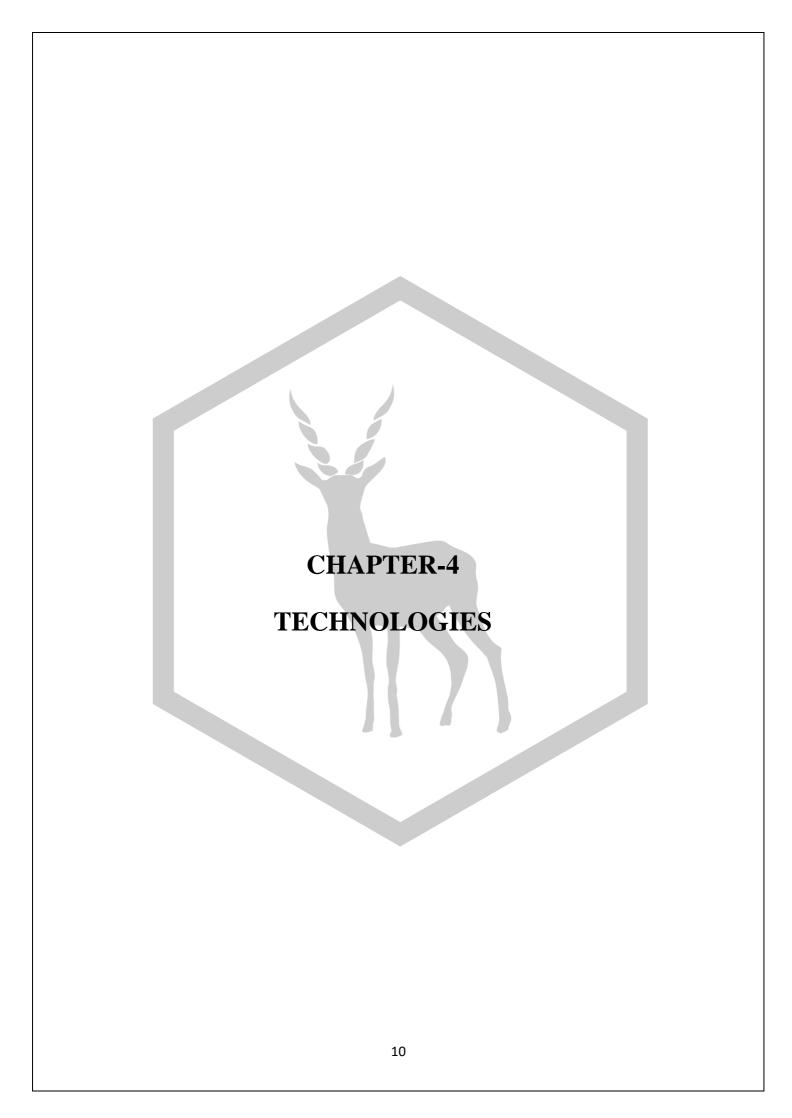
SOFTWARE REQURIMENTS SPECIFICATIONS

Software Requirements:

- The software and hardware requirements for sentiment analysis of Amazon Alexa reviews may vary depending on the specific system and implementation.
- Software Requirements:
- Programming Language: Python, Java, or any other language suitable for machine learning and natural language processing.
- Text Processing Libraries: NLTK, Spacy, Gensim, etc.
- Machine Learning Frameworks: TensorFlow, PyTorch, Scikit-Learn, etc.
- Sentiment Analysis Libraries: TextBlob, VADER, etc.
- Database Management System: MySQL, MongoDB, or any other suitable database.

Hardware Requirements:

- Processor: At least a quad-core processor (Intel i5 or equivalent).
- RAM: At least 8GB RAM or higher.
- Storage: At least 100GB of storage capacity.
- GPU: A high-end GPU like NVIDIA RTX 3090 or AMD Radeon RX 6800 XT may be required for large-scale training of machine learning models..
- It is important to note that the software and hardware requirements may change based on the size of the dataset, the complexity of the machine learning models, and the level of accuracy required in the sentiment analysis.



TECHNOLOGIES

- Natural Language Processing (NLP): A technology that helps computers
 understand and process human language, which can be used to analyze Amazon
 Alexa reviews.
- **Sentiment Analysis Models:** Machine learning algorithms that can classify text data as positive, negative, or neutral based on the sentiment expressed in the text.
- **Data Storage and Processing Technologies:** Technologies that help handle large volumes of data, such as Amazon S3, Apache Hadoop, or Apache Spark.
- **Visualization Tools:** Tools that help create interactive dashboards and visualizations to display the results of sentiment analysis, such as Tableau or D3.js.
- Cloud Services: Services that provide scalability, security, and reliability, such as Amazon Web Services (AWS) or Microsoft Azure.
- **Programming Languages and Frameworks:** Tools used to develop a sentiment analysis system, such as Python, R, TensorFlow, and Scikit-learn.

Modules involved in the project :

- Data Collection: The first step is to collect data from various sources, such as Amazon's customer reviews, product descriptions, and other relevant data sources.
- Preprocessing: The collected data is preprocessed to remove any unwanted information such as stop words, punctuations, special characters, and numbers. This process helps to reduce the size of the data and makes it more manageable.
- Sentiment Analysis: This is the core module that performs sentiment analysis on the
 preprocessed data. The module uses Natural Language Processing (NLP) techniques
 such as tokenization, part-of-speech tagging, and sentiment lexicons to analyze the
 text and identify the sentiment of the reviews.
- Machine Learning: Machine learning algorithms are used to classify the sentiment of
 the reviews as positive, negative, or neutral. This module uses a training dataset to
 train the model and then uses the trained model to classify new reviews.
- Output Generation: The final module generates the output of the sentiment analysis, which can be displayed in the form of visualizations or text summaries. The output

provides insights into the overall sentiment of the reviews and helps businesses make data-driven decisions.

- pandas: Used for data manipulation and analysis, specifically for loading and cleaning the data in this case.
- numpy: Used for numerical computation, specifically for creating arrays and matrices.
- nltk: A natural language processing library in Python used for tokenizing, stemming, and lemmatizing the text data.
- sklearn: A machine learning library in Python used for feature extraction, data splitting, and model training and evaluation.
- TfidfVectorizer: A feature extraction technique used to transform text data into a matrix of TF-IDF features.
- train_test_split: A function from the sklearn library used to split the data into training and testing sets.
- classification_report: A function from the sklearn library used to evaluate the performance of the sentimental analysis model.
- SVC: A Support Vector Machine algorithm from the sklearn library used for training the sentimental analysis model.



CODING:

1. Loading the dataset: Load the data and import the libraries.

2. Data cleaning:

- Deleting redundant columns.
- Renaming the columns
- Dropping duplicates
- Cleaning individual columns
- Removing the NAN values in dataset

3. Recommendation System

Import libraries and datasets

import pandas as pd

import numpy as np

import re

import sklearn as sk

import seaborn as sns

from matplotlib import pyplot as plt

plt.style.use('ggplot')

import os

from wordcloud import WordCloud

import string

import nltk

nltk.download('stopwords')

from nltk.corpus import stopwords

from sklearn.model_selection import train_test_split

from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer

from sklearn.svm import SVC, LinearSVC

from sklearn.naive_bayes import MultinomialNB

```
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import classification_report, confusion_matrix
[nltk data] Downloading package stopwords to /usr/share/nltk data...
[nltk data]
                  Package stopwords is already up-to-date!
df = pd.read_csv("../input/amazon-alexa-reviews/amazon_alexa.tsv",sep=\t')
df.head()
Exploratory data analysis
df.describe()
df['length'] = df['verified_reviews'].apply(len)
df.describe()
df['rating'].value_counts()
# Plot the count plot for the ratings
plt.figure(figsize=(6, 2))
sns.countplot( y= df['rating'])
plt.title("Number of reviews per ratings")
df['feedback'].value_counts()
plt.figure(figsize=(6, 2))
sns.countplot(y=df['feedback'] )
plt.title("Number of reviews per categories")
# Plot the histogram of the length column
df['length'].plot(bins=100, kind='hist')
Plot word-cloud
list_cat = df['feedback'].unique()
print(list_cat)
word_list_cat = []
for cat in list cat:
  val = df[df['feedback'] == cat]
  sentences = val['verified_reviews'].tolist()
  sentences_as_one_string =" ".join(sentences)
word_list_cat.append(sentences_as_one_string)
```

```
print(len(word_list_cat))
Word-cloud of positive reviews
plt.figure(figsize=(8,8))
plt.imshow(WordCloud().generate(word_list_cat[0]))
plt.title("Positive reviews")
Word-cloud of negative reviews
plt.figure(figsize=(8,8))
plt.imshow(WordCloud().generate(word_list_cat[1]))
plt.title("Negative reviews")
Train and test
def message cleaning(message):
  Test_punc_removed = [char for char in message if char not in string.punctuation]
  Test_punc_removed_join = ".join(Test_punc_removed)
  Test_punc_removed_join_clean = [word for word in Test_punc_removed_join.split() if
word.lower() not in stopwords.words('english')]
  return Test_punc_removed_join_clean
vectorizer = CountVectorizer(analyzer = message_cleaning, dtype = np.uint8)
df_countvectorizer = vectorizer.fit_transform(df['verified_reviews'])
print(df_countvectorizer.shape)
X = pd.DataFrame(df\_countvectorizer.toarray())
y = df['feedback'].values
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
Linear Support vector machine
SVC_classifier = LinearSVC()
SVC_classifier.fit(X_train, y_train)
print(SVC_classifier.score(X_test, y_test))
# Predicting the Test set results
y_predict_test = SVC_classifier.predict(X_test)
cm = confusion_matrix(y_test, y_predict_test)
sns.heatmap(cm, annot=True)
print(classification_report(y_test, y_predict_test))
```



Prints first five rows:

	rating	date	variation	verified_reviews	feedback
0	5	31-Jul-18	Charcoal Fabric	Love my Echo!	1
1	5	31-Jul-18	Charcoal Fabric	Loved it!	1
2	4	31-Jul-18	Walnut Finish	Sometimes while playing a game, you can answer	1
3	5	31-Jul-18	Charcoal Fabric	I have had a lot of fun with this thing. My 4	1
4	5	31-Jul-18	Charcoal Fabric	Music	1

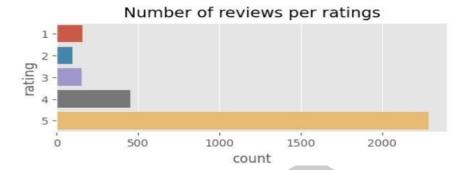
Describe

	rating	feedback
count	3150.000000	3150.000000
mean	4.463175	0.918413
std	1.068506	0.273778
min	1.000000	0.000000
25%	4.000000	1.000000
50%	5.000000	1.000000
75%	5.000000	1.000000
max	5.000000	1.000000

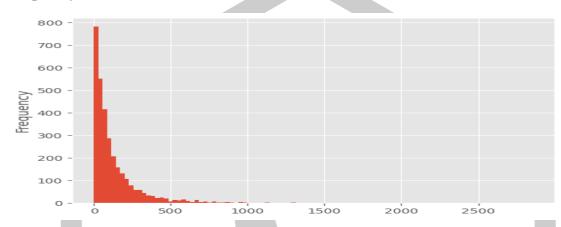
Calculating length

	rating	feedback	length
count	3150.000000	3150.000000	3150.000000
mean	4.463175	0.918413	132.049524
std	1.068506	0.273778	182.099952
min	1.000000	0.000000	1.000000
25%	4.000000	1.000000	30.000000
50%	5.000000	1.000000	74.000000
75%	5.000000	1.000000	165.000000
max	5.000000	1.000000	2851.000000

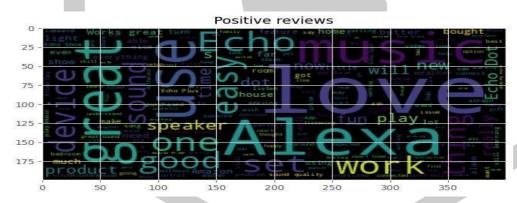
Rating



Frequency

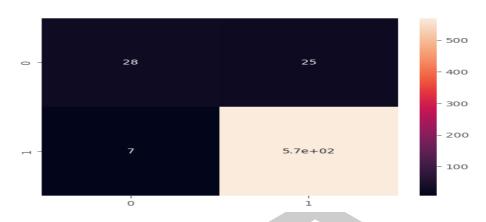


Positive reviews



Negative reviews





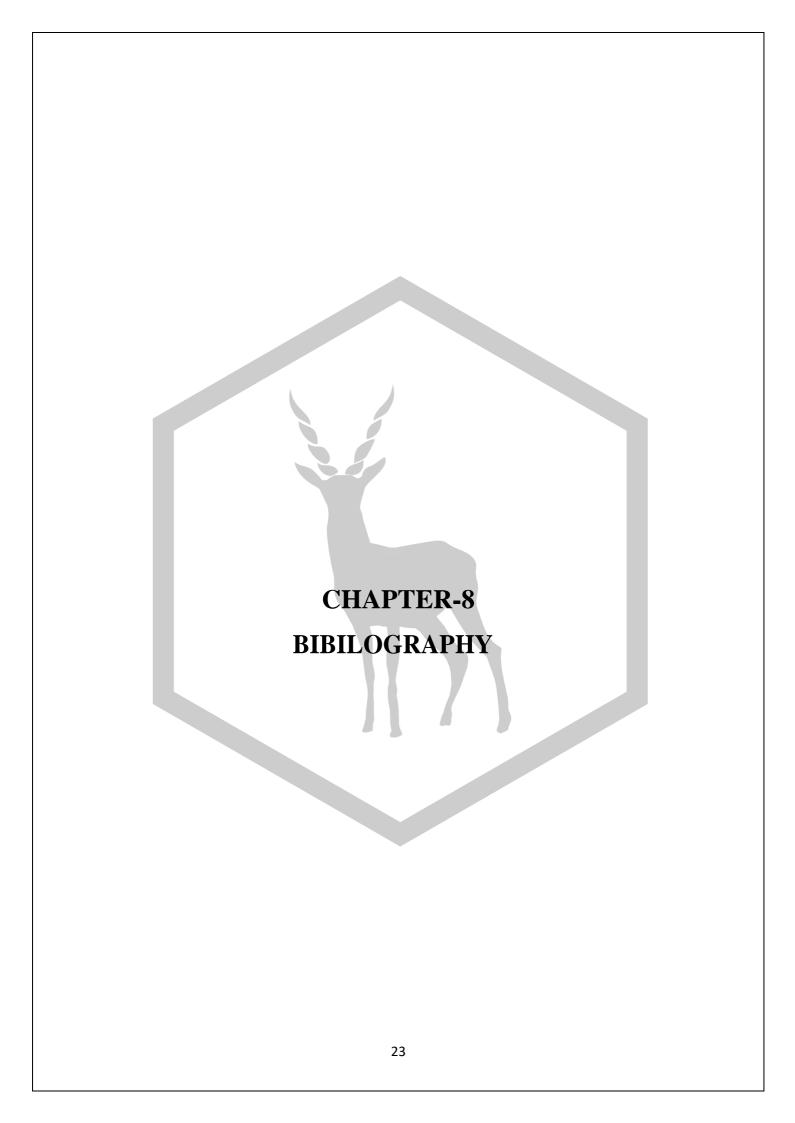
Accuracy

support	f1-score	recall	precision	
53	0.64	0.53	0.80	0
577	0.97	0.99	0.96	1
630	0.95			accuracy
630	0.80	0.76	0.88	macro avg
630	0.94	0.95	0.94	weighted avg



CONCLUSION

In conclusion, Amazon Alexa reviews sentimental analysis is an important process for businesses to gain insights into the sentiment of their customers towards their products or services. The process involves several modules such as data collection, preprocessing, sentiment analysis, machine learning, and output generation. By analyzing the sentiment of the reviews, businesses can identify areas for improvement and make data-driven decisions to enhance the customer experience. Additionally, sentiment analysis can also help businesses identify positive reviews and leverage them for marketing purposes. Overall, Amazon Alexa reviews sentimental analysis is a valuable tool for businesses to understand customer sentiment and improve their products and services.



BIBILOGRAPHY:

- **1.** Purohit, H., & Sheth, A. (2016). Sentiment analysis of twitter data: A survey of techniques. ACM Computing Surveys (CSUR), 49(2), 28.
- **2.** Zhang, Y., Wallace, B., & Liu, J. (2019). Rationale-augmented convolutional neural networks for text classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 796-805).
- **3.** Chen, L., & Ibrahim, R. (2018). Sentiment analysis using deep learning: A review. arXiv preprint arXiv:1801.07883.
- **4.** Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international AAAI conference on weblogs and social media.
- **5.** Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
- **6.** Zhang, D., & Lee, W. S. (2006). Automatic keyword extraction for sentiment analysis. In Proceedings of the 20th international conference on computational linguistics (pp. 787-794).
- **7.**Luo, J., Wu, J., & Yang, X. (2019). A deep learning framework for aspect-based sentiment analysis. Information Sciences, 495, 24-36.
- **8.** Amazon Developer Documentation, "Alexa Reviews API." Available online: https://developer.amazon.com/en-US/docs/alexa/custom-skills/reviews-api.html (accessed on May 12, 2023).

