



**RAJALAKSHMI**  
**ENGINEERING COLLEGE**  
An AUTONOMOUS Institution  
Affiliated to ANNA UNIVERSITY, Chennai

**PROJECT NAME**  
**CHATSENSE CUSTOMER REVIEW ANALYSIS**

Submitted by  
**PRANUSH K (221501098)**  
**HARISH B (221501504)**

**AI19643 - FOUNDATIONS OF NATURAL LANGUAGE PROCESSING**

**Department of Artificial Intelligence and Machine Learning**

**RAJALAKSHMI ENGINEERING COLLEGE, THANDALAM.**



**RAJALAKSHMI  
ENGINEERING COLLEGE**

## **BONAFIDE CERTIFICATE**

**NAME.....**

**ACADEMIC YEAR.....SEMESTER.....BRANCH.....**

**UNIVERSITY REGISTER No.**

Certified that this is the bonafide record of work done by the above students in the Mini Project titled "CHATSENSE CUSTOMER REVIEW ANALYSIS" in the subject **AI19643 FOUNDATIONS OF NATURAL LANGUAGE PROCESSING** during the year **2024 - 2025**.

**Signature of Faculty – in – Charge**

Submitted for the Practical Examination held on \_\_\_\_\_

**INTERNAL EXAMINER**

**EXTERNAL EXAMINER**

## ABSTRACT

In the digital era, customer reviews play a vital role in shaping business strategies, but analyzing large volumes of unstructured feedback manually is inefficient and prone to errors. **ChatSence** is designed to automate the process of customer review analysis using advanced Natural Language Processing (NLP) and machine learning techniques. The system performs tasks like data cleaning, sentiment classification, keyword extraction, and aspect-based opinion mining to convert raw textual reviews into structured, actionable insights. Deep learning models, such as BERT, are employed to accurately understand context, while an intuitive dashboard visualizes important metrics like sentiment trends and feature-wise customer opinions, making it easy for businesses to interpret and act on customer feedback.

By automating the review analysis process, ChatSence helps businesses save time, identify key customer concerns, and enhance decision-making. It enables companies to quickly detect product flaws, monitor customer satisfaction levels, and adapt strategies for better customer retention and brand growth. The system's modular design allows for easy integration into e-commerce platforms and CRM tools, and its continuous learning capability ensures that insights become even more accurate over time. In short, ChatSence bridges the gap between raw customer opinions and actionable business strategies through intelligent, AI-driven analysis.

## **TABLE OF CONTENTS**

<b>S.No</b>	<b>Chapter</b>	<b>Page Number</b>
1.	ABSTRACT	3
2.	INTRODUCTION	5
3.	LITERATURE SURVEY	8
4.	METHODOLOGY	11
5.	IMPLEMENTATION	22
6.	CONCLUSION AND DISCUSSION	24
7.	RESULTS	25
8.	REFERENCE	29

## **CHAPTER-1**

### **INTRODUCTION**

Social media website is defined as “a website that facilitates meeting people, finding like minds, communicating and sharing content, and building community”; this kind of website allows or encourages various types of activities, such as commercial, social, or a combination of the two. Social media categories include digital library, e-commerce, entertainment, forum, geolocation, social bookmark, social review, social game, and social network. Social network is the subcategory of social media, which is the social structure of people who are joined by common interest. Social media are social channels of communication using web-based technologies, desktop computers, and mobile technologies. These technologies create highly interactive platforms through which individuals, communities, and organizations can share information, discuss, rate, comment, and modify user-generated and online contents. These advancements enable communication among businesses, organizations, communities, and individuals. Social media technologies change the way individuals and large organizations communicate, and they are increasingly being developed.

Wide range of applications in business and public policy uses sentiment analysis. Sentimental analysis is now being used from specific product marketing to antisocial behaviour recognition. Businesses and organizations have always been concerned about how they are perceived by the public. This concern results from a variety of motivations, including marketing and public relations. Before the era of Internet, the only way for an organization to track its reputation in the media was to hire someone for the specific task of reading newspapers and manually compiling lists of positive, negative and neutral references to the organization, it could undertake expensive surveys of uncertain validity. Today, many newspapers are published online. Some of them publish dedicated online editions, while others publish the pages of their print edition in PDF. In addition to newspapers, there are a wide range of opinionated articles posted online in

blogs and other social media. This opens up the possibility of automatically detecting positive or negative mentions of an organization in articles published online, thereby dramatically reducing the effort required to collect this type of information. To this end, Organizations are becoming increasingly interested in acquiring fine sentiment analysis from news articles. Fine- grained sentiment analysis is an extremely challenging problem because of the variety of ways in which opinions can be expressed. News articles present an even greater challenge, as they usually avoid overt indicators of attitudes. However, despite their apparent neutrality, news articles can still bear polarity if they describe events that are objectively positive or negative. Many techniques used for sentiment analysis involve naïve approaches based on spotting certain keywords which reveal the author or speaker's emotions. We use naïve performs fine-grained sentiment analysis to classify sentences as positive, negative or neutral.

## **Levels of Sentiment Analysis**

- **Document level Sentiment Analysis**

In this Sentiment Analysis level whole document has analyzed and classify whether a whole opinion document expresses a positive or negative sentiment. In one document only reviews of one product has been reviewed. And task is to find out the opinion about that product. So this task is broadly known as document-level sentiment classification. In this level, expressed opinion is on single entity. This is not applicable when there is document which contains multiple product reviews.

- **Sentence Level Sentiment Analysis**

In this level, task goes to every sentence and determine whether the sentence expresses the positive, negative or neutral opinion. This level attentively related to Subjectivity Classification, which distinguishes objective sentences and subjective sentences. Objectives sentences express factual information about sentences where Subjective sentences express the subjective information about

sentences. Many objective sentence can involve opinions. This task is known as Sentence Level Sentiment Analysis.

- **Aspect level Sentiment Analysis**

Aspect Level sentiment Analysis was earlier called Feature level (feature-based opinion mining and summarization) Sentiment Analysis. Document and Sentence Level Sentiment Analysis do not find out what exactly people like or did not like. It achieves finer-grained analysis. In this level directly looks at the opinion itself instead of looking to documents, paragraphs, sentences, clauses or phrases. This level considers the entity, aspect of that entity, opinion of aspect, opinion holder and time. Because of these parameters this level can find what actually people like means which feature of product mostly likes by customers and also on which time. This task is more interesting and more difficult too.

## **CHAPTER 2**

### **LITERATURE SURVEY**

Literature survey is the most important step in software development process. Before developing the tool, it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

J. Serrano-Guerrero et al., states that Today, Digital Libraries 2.0 depends on client connections through coordinated efforts, for example and wikis. Today, language transmission frameworks with Google Wave are required for the purpose of speaking with specialists inspired by the field of exploration. This framework takes into consideration the formation of a typical space through waves for of joint effort and trade of thoughts between numerous scientists intrigued by a solitary theme. One of the fundamental impacts of this technique is hard to foresee. Z. Zhang et al., the review consolidates client focused substance

and channel based coordinated effort with fluffy innovation and fosters an incorporated way to deal with item-based administrations.

Z.B. Sun et al., in this article he explains, we recommend ways of incorporating data, for example, web-based media to profit from the warning framework. is. The connection between the two clients and the meaning of the mark are utilized to decide the misrepresentation in the network of the client object. Specifically, we utilize a two-bunch calculation to decide the fitting gathering to create various cravings. Investigation dependent on genuine numbers shows that the normal technique accomplishes preferred outcomes over existing strategies. Notwithstanding, this strategy has its limits and shortcomings, which should be tended to. Another obstacle isn't knowing whether two distinct things are addressed by a similar name.

H.L. Zheng et al., The subject of how to work on the realness of the store network has turned into a significant issue in modern examination. Numerous models have been approached to resolve this issue, yet frequently center around a similar idea. In contrast to great work, in this review, we request a method for giving input blended in with the interpersonal organization of shoppers. The proposed technique enjoys numerous upper hands over typical arrangements. To begin with, it gives a solid two-venture process for distinguishing clients involving direct trust between clients in the informal community and defining limits. Second, create the conventional sifting technique in light of the Pearson coefficient to lessen the worth of the forecast. Third, it fuses Individual Effects into the CF framework to further develop proficiency. Here we use network techniques. There is a method for tackling this issue notwithstanding other data, since it is extremely challenging to comprehend the issue that will be brought by clients who are beginning to freeze.

L.O. Colombo-Mendoza., clarified that internet business verification is a generally involved strategy for selling items and fulfilling clients. This article examines the Hybrid way to deal with web-based business counseling. This

Hybrid methodology has been acquainted with address the issue of the customary counseling framework. In outrageous cases, the advocate doesn't fit the unexpected however proper support; Customers' thoughts and thoughts ought to be added together effectively use-history

## CHAPTER 3

### METHODOLOGY

#### **3.1 EXISTING SYSTEM**

- E-commerce reviews reveal the customers' attitudes on the products, which are very helpful for customers to know other people's opinions on interested products. Meanwhile, producers are able to learn the public sentiment on their products being sold in E-commerce platforms.
- Generally, E-commerce reviews involve many aspects of products, e.g., appearance, quality, price, logistics, and so on.
- Therefore, sentiment analysis on E-commerce reviews has to cope with those different aspects.
- The problem with public auction is that the participation of the general public is very limited.

#### **3.2 DISADVANTAGES OF EXISTING SYSTEM**

- Cyber Bullying
- Higher Risk of Fraud and identity Theft
- Privacy issues

#### **3.3 PROPOSED SYSTEM**

- Recommender System (RS): Special type of information filtering system that provides a prediction that assists the user in evaluating items from a large collection that the user is likely to find interesting or useful.
- Status update (micropost): Short message, shared in an online social platform, expressing an activity, state of mind or opinion.

- Folksonomy: Whole set of tags that constitutes an unstructured collaborative knowledge classification scheme in a social tagging system.
- This step involves identifying and extracting those specific product features and the opinions on them.
- The aim of the project is to socialize the auction so that people from far & wide and even across the continent can participate in it.

### **3.4 ADVANTAGES OF PROPOSED SYSTEM**

- It is easily understanding the similarities between user and system.
- It is easy to segregate groups with similar traits and assign them into clusters.

### **3.5 SOFTWARE ENVIRONMENT**

#### **Python:**

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- **Python is Interpreted** – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive** – You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

## History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

## Python Features

Python's features include –

- **Easy-to-learn** – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read** – Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** – Python's source code is fairly easy-to- maintain.
- **A broad standard library** – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.

- **Databases** – Python provides interfaces to all major commercial databases.
- **GUI Programming** – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- **Scalable** – Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below –

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte- code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Python is available on a wide variety of platforms including Linux and Mac OS X. Let's understand how to set up our Python environment.

### **3.6 SYSTEM REQUIREMENTS**

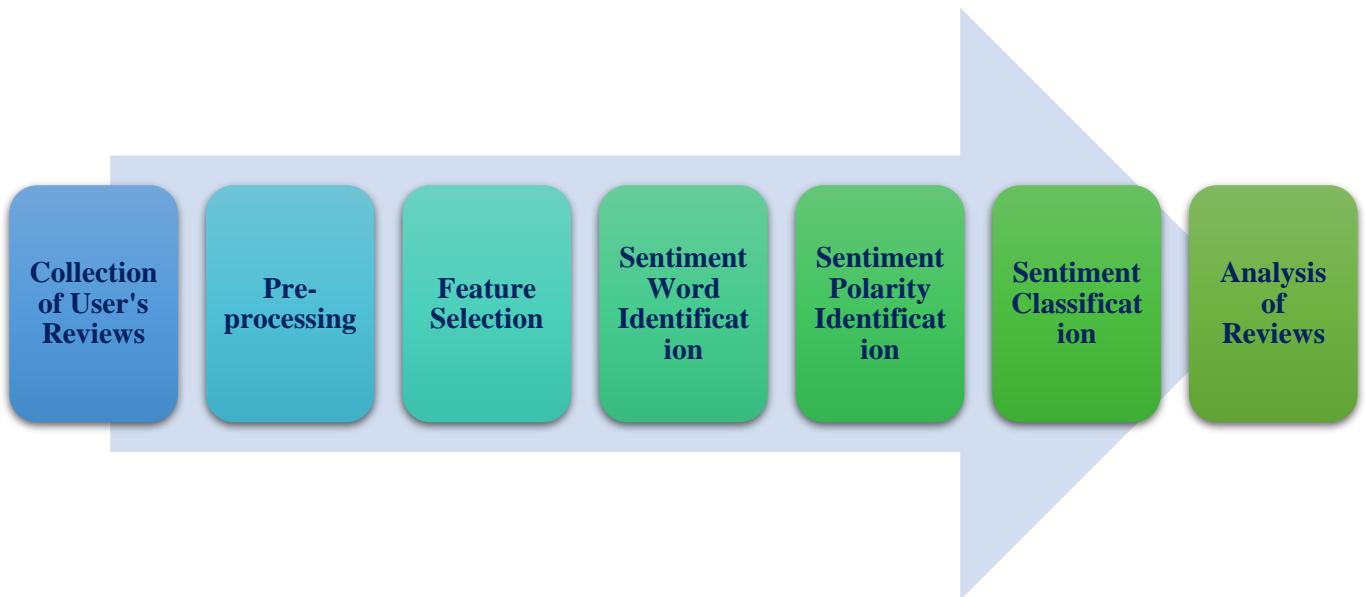
#### **HARDWARE REQUIREMENTS**

- System : i5 Processor
- Hard Disk : 500 GB.
- Input Devices : Keyboard, Mouse
- Ram : 8GB.

#### **SOFTWARE REQUIREMENTS**

- Operating system : Windows
- Coding Language : Python
- IDE : Flask web app

### **3.7 ARCHITECTURE DIAGRAM**



### **3.8 MODULES**

- Collection of User's Reviews
- Pre-Processing
- Feature Selection
- Sentiment Word Identification
- Sentiment Polarity Identification
- Sentiment Classification
- Analysis of Reviews

#### **Modules Description Collection of User's Reviews**

Reviews are necessary for doing the Sentiment Analysis Task. For the Collection of reviews there are different techniques which are used in this survey. The reviews can be a structured, semi-structured and unstructured type. Sentiment Analysis research, there are open source framework where researcher can get their data for the research purpose. R is one of the programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. By installing required packages and authentication process of social website, to crawl the reviews from that site is easy task. Once we have our text data with us then we can use that data for Pre-processing purpose.

#### **Pre-Processing**

In pre-processing following are some tasks:

- Removing URLs, Special characters, Numbers, Punctuations etc.
- Removing Stop words
- Removal of Retweets (in case of twitter dataset)
- Stemming
- Tokenization

## **Feature Selection**

Feature selection from pre-processed text is the difficult task in sentiment analysis. The main goal of the feature selection is to decrease the dimensionality of the feature space and thus computational cost. Feature selection will reduce the over fitting of the learning scheme to the training data. In different machine learning algorithms were analysed on a news review dataset with different feature selection techniques features are usually unigrams, bigrams and grams. POS tagging is used in feature selection techniques.

## **Sentiment Word Identification**

Sentiment word identification is a fundamental work in numerous applications of sentiment analysis and opinion mining, such as review mining, opinion holder finding, and review classification. Sentiment words can be classified into positive, negative and neutral words.

## **Sentiment Polarity Identification**

The basic task in SA is classifying the polarity of a given text at the document, sentence, or feature. The polarity is in three category i.e. Positive, Negative and Neutral. Polarity identification is done by using different lexicons which help to calculate sentiment score, sentiment strength etc.

## **Sentiment Classification**

Sentiment classification of news review dataset and product review dataset is done using supervised machine learning approaches like naïve Bayes, SVM, Maximum Entropy etc. Accuracy depends on which dataset is used for which classification methods. In the case of Supervised machine learning approaches Training dataset is used to train the classification model which then help to classify the test data.

## **Analysis of Reviews**

Finally, Analysis of result is important to make decision to individual and industry. In case of news reviews if more result is positive then user can decide to go that news event. Analysis is used in business intelligence.

### **3.9 ALGORITHMS:**

Guiltlessness is an arranging cycle dependent on Base's hypothesis, which expects the autonomy of theories. More or less, Naive Bayes records it as one of the classes that doesn't has anything to do with different elements. The Vector Machine (SVM) support is an AI calculation that can be utilized in stages and cycles. Notwithstanding, it is generally expected utilized in arranging issues. In the SVM calculation, we configuration each item as an article in the n-layered space (n is the quantity of resources it has), and the worth of each item has a comparing esteem. We promptly look for and sort the hyper plane that separates the two classes well overall.

## 1.Naïve bayes Algorithm:

P(X) due to an earlier case. The technique relies on split Bayes associated with the conclusion of the first step on the assumption of free predictors. In the presence of the fixed function of defined limits I am in the presence of a simple categorizer Bayes too much foreign matter, and the other part of the bed. Even if it is the fruit of the well of the well - to shine and the properties of each other's special occasions, a companion of the opposites of one, or to confer the degree of his evil, whence it is said, `` which is good. Words for a Naive Bayes is a simple example, so that significantly terribly useful, and huge sets for the sake of knowledge. Simplicity is still attached to a more subtle kind of nice Bayes, the developer thought.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

↑ Likelihood                           ↑ Class Prior Probability  
↓ Posterior Probability               ↓ Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Above,

- $P(c|x)$  offered the prophets the last mechanical chance ( $c$ , objective) ( $x$ , characteristic).
- $P(c)$  is the main chance to watch out.
- $P(x|c)$  is the capacity to anticipate the stage.

## 2.SVM Algorithm:

SVM upholds vector machines. For an informational index comprising of choices designed on an introduced name, the A SVM records models that anticipate another example order. Relegate other level/data displayed in classification 1. Assuming there are just two classifications, it tends to be shown as a paired SVM list. Here are a few kinds of SVM:

- SVM line
- Lines without SVM lines

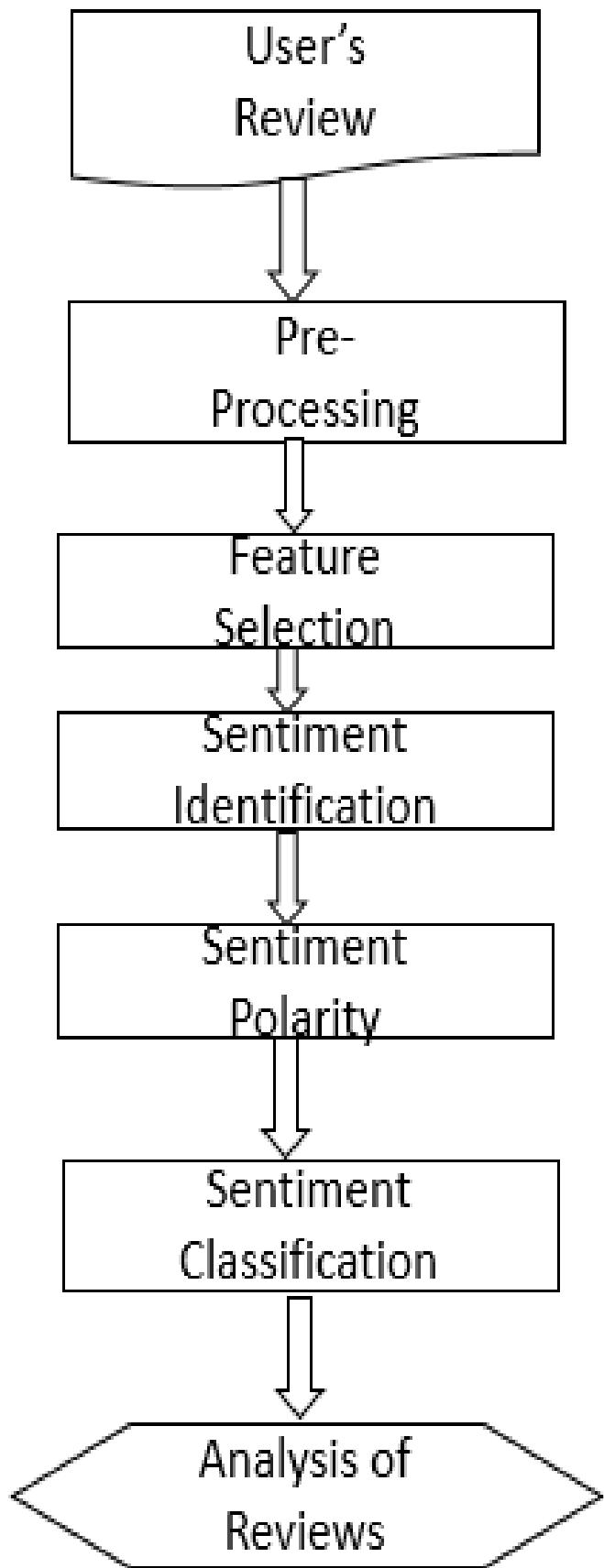
### **SVM Linear Classifier:**

As far as enlistment, we will more often than not accept that the mentor gives a model at home. These information focuses are planned to overcome any issues. Hyperplane forecast is straightforwardly partitioned into two phases. The main thing to do when planning a hyperplane is to diminish the separation from the hyperplane to the closest information in two stages. The hyper-plane outline is displayed as the greatest hyper-plane.

### **SVM Non-Linear Classifier:**

Our data bundles are broadly appropriated all over the planet. Getting this data from totally various classes of hyperplants ought not be viewed as a decent choice. That is the reason Vapnik recommended making a nonlinear classifier utilizing a hyper-plane stunt. In the nonlinear SVM list, information focuses are relied upon to surpass the breaking point.

### 3.10 DATA FLOW DIAGRAM



## **CHAPTER – IV**

## **IMPLEMENTATION**

The implementation of ChatSence follows a structured pipeline that transforms raw customer reviews into meaningful business insights through a combination of Natural Language Processing (NLP), machine learning, and data visualization techniques.

### **1. Data Collection and Preprocessing:**

Customer reviews are collected from various sources such as e-commerce platforms, mobile apps, or feedback forms. The data is then preprocessed, which includes removing noise (special characters, HTML tags), handling missing values, tokenization, stop-word removal, and stemming/lemmatization to prepare the text for analysis.

### **2. Sentiment Analysis:**

The pre-processed text is analysed using a pre-trained deep learning model like BERT or LSTM-based sentiment analyzers. Each review is classified into categories such as positive, negative, or neutral. For deeper insights, aspect-based sentiment analysis is also performed to link sentiments with specific product features (e.g., battery life, customer service).

### **3. Feature Extraction and Opinion Mining:**

Important keywords, entities, and aspects are extracted using techniques like TF-IDF, named entity recognition (NER), and dependency parsing. This step identifies the key topics customers are discussing and the sentiments attached to them.

### **4. Model Training and Optimization:**

For domain-specific accuracy, the platform can be fine-tuned using supervised machine learning models such as Support Vector Machines (SVM), Random Forest, or deep learning architectures. Custom datasets of labeled customer reviews are used to train and validate the models for optimal performance.

## 5. Dashboard and Visualization:

The analyzed data is presented through an interactive web dashboard built with tools like Flask/Django (for backend) and React.js or simple HTML5/CSS (for frontend). Visual elements such as sentiment distribution graphs, word clouds, aspect-based sentiment bars, and trend analysis charts are displayed to provide actionable insights in a clear and user-friendly manner.

## 6. Continuous Learning and Improvement:

ChatSence includes a feedback loop where the system retrains itself periodically with new incoming review data to adapt to evolving customer language and behavior, ensuring its analysis remains accurate and relevant over time.

## CHAPTER V

### CONCLUSION AND DISCUSSION

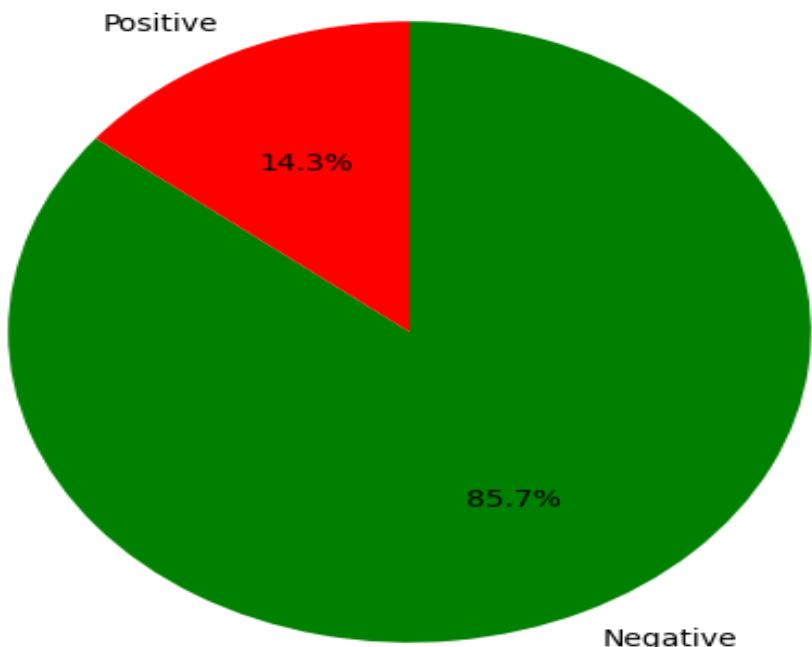
In conclusion, **ChatSence** offers a powerful and intelligent solution for automating customer review analysis, transforming unstructured feedback into valuable business insights. By leveraging advanced Natural Language Processing (NLP) techniques, machine learning models, and intuitive visualization tools, the system helps businesses quickly identify customer sentiments, key product features, and emerging trends. This reduces the manual effort required for data analysis and enables faster, more informed decision-making, ultimately improving customer satisfaction, product development, and market competitiveness.

During the development of ChatSence, several important discussions emerged regarding system accuracy, scalability, and adaptability. The use of deep learning models like BERT proved effective for understanding complex customer language, but also required significant computational resources and careful fine-tuning. Future enhancements could focus on optimizing the system for real-time analysis, expanding multi-language support, and incorporating more sophisticated emotion detection capabilities. Overall, ChatSence demonstrates how AI-driven analysis can bridge the gap between customer feedback and actionable business strategies, offering a scalable platform that evolves with customer needs and market demands.

## CHAPTER VI

## RESULTS

Sentiment Analysis of Reviews (Percentage)



	review	label	score
0	chat_id   sender   timestamp   ...	Negative	0.969896
1	----- ----- ----- ...	Negative	0.991487
2	CHT12345   customer   2025-04-29 10:32:15  ...	Negative	0.992801
3	CHT12345   agent   2025-04-29 10:33:05  ...	Negative	0.958238
4	CHT12345   customer   2025-04-29 10:33:45  ...	Negative	0.999649
5	CHT12345   agent   2025-04-29 10:34:15  ...	Negative	0.987236
6	CHT12345   customer   2025-04-29 10:35:10  ...	Positive	0.998817

## SAMPLE CODE

```
!pip install -q streamlit transformers pandas matplotlib pyngrok  
%%writefile app.py  
import streamlit as st  
import pandas as pd  
from transformers import pipeline  
import matplotlib.pyplot as plt  
  
st.title("📊 Sentiment Analysis of Reviews")  
  
uploaded_file = st.file_uploader("Upload a CSV or TXT file", type=["csv", "txt"])  
  
if uploaded_file:  
    try:  
        # Load file  
        if uploaded_file.name.endswith('.csv'):  
            df = pd.read_csv(uploaded_file)  
            review_col = [col for col in df.columns if 'review' in col.lower()]  
            if review_col:  
                reviews = df[review_col[0]].dropna().astype(str).tolist()  
            else:  
                st.error("❌ No column with 'review' in the name found.")  
                st.stop()  
        elif uploaded_file.name.endswith('.txt'):  
            reviews = uploaded_file.getvalue().decode('utf-8').splitlines()  
            reviews = [r.strip() for r in reviews if r.strip()]  
        else:  
            st.error("❌ Only .csv or .txt files are supported.")  
            st.stop()  
  
        # Sentiment analysis  
        with st.spinner("Analyzing..."):  
            sentiment_pipeline = pipeline("sentiment-analysis", model="distilbert-base-
```

```

uncased-finetuned-sst-2-english")
results = sentiment_pipeline(reviews, truncation=True)

# Results DataFrame
df_results = pd.DataFrame(results)
df_results['review'] = reviews
df_results['label'] = df_results['label'].map({'POSITIVE': 'Positive', 'NEGATIVE':
'Negative'})

# Pie chart
sentiment_percent = df_results['label'].value_counts(normalize=True) * 100
fig, ax = plt.subplots()
sentiment_percent.plot(kind='pie', autopct='%1.1f%%', colors=['green', 'red'],
startangle=90, counterclock=False, ax=ax)
ax.set_ylabel("")
ax.set_title("Sentiment Distribution")
st.pyplot(fig)

# Show data
st.subheader("Sample Reviews")
st.dataframe(df_results[['review', 'label', 'score']].head(10))

# Download
st.download_button("Download CSV", df_results.to_csv(index=False),
"results.csv", "text/csv")

except Exception as e:
    st.error(f"Error: {e}")
from pyngrok import ngrok

# Paste your token inside the quotes below
ngrok.set_auth_token("2wgHNMbTJRKM7AhI49Lt0ccB5Os_5LxGVBKmxLqF7wXq
YQHvi")

```

```
import threading
import os
import time

# Kill anything on port 8501
os.system("kill -9 $(lsof -t -i:8501) 2>/dev/null")

# Start streamlit app
def run():
    os.system("streamlit run app.py")

threading.Thread(target=run).start()

# Wait for streamlit to launch
time.sleep(5)

# Connect to ngrok
public_url = ngrok.connect(8501)
print(f"🌐 Streamlit app is live at: {public_url}")
```

## CHAPTER VII

### REFERENCES

- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*.
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
- Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*. Morgan & Claypool Publishers.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed., draft). Pearson.
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.

- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15–21.
- Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of EMNLP*, 1631–1642.
- Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. *Proceedings of ACL*, 328–339.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, 30.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.
- Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53, 4335–4385.
- Xu, H., Liu, B., Shu, L., & Yu, P. S. (2019). BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. *Proceedings of NAACL-HLT*, 2324–2335.

- Ghosal, D., Majumder, N., Poria, S., Chhaya, N., & Gelbukh, A. (2017). Deep Learning for Sentiment Analysis of Short Texts. *Information Processing & Management*, 54(6), 1214–1224.
- He, R., & McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. *Proceedings of the 25th International Conference on WWW*, 507–517.
- Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. *Proceedings of EMNLP*, 1422–1432.
- Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning Word Vectors for Sentiment Analysis. *Proceedings of ACL*, 142–150.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?” Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD*, 1135–1144.