Sentiment Analysis of Customer Reviews using AI-based Techniques for E-commerce Websites

A MINOR PROJECT REPORT

Submitted by

RISHI BHATT [RA2011030010185]

ASHUTOSH PAIKARAY [RA2011030010196]

HARSHIT KRISHNA [RA2011030010201]

Under the guidance of

Dr.K.A Varun Kumar

Assistant Professor, Department of Networking and Communications

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

S.R.M.Nagar, Kattankulathur, Chengalpattu District

MAY 2023

SRM INSTITUTION OF SCIENCE AND TECHNOLOGYKATTANKULATHUR-603203

BONAFIDE CERTIFICATE

Certified that this lab report titled **"Sentiment Analysis of Customer Reviews using AI-based Techniques for E-commerce Websites"** is the bonafide work done by **Rishi Bhatt (RA2011030010185), Ashutosh Paikaray (RA2011030010196), Harshit Krishna (RA2011030010201)** III Year/VI Sem B.Tech(CSE) who carried out the mini project work under my supervision for the course 18CSC305J - Artificial Intelligence in SRM Institute of Science and Technology during the academic year 2022-2023 (Even Sem). Certified further, that to the best of my knowledge the work reported herein does not form part of any other work.

SIGNATURE

Dr.K.A Varun Kumar

AI - Course Faculty

Assistant Professor

Department of NWC

ABSTRACT

This project aims to develop an Artificial Intelligence-based system for sentiment analysis of customer reviews for e-commerce websites. The system will analyze the text content of the reviews and classify them as positive, negative, or neutral. The project will involve data collection, preprocessing, text embedding, model selection, and deployment. The team will collect customer reviews from e-commerce websites and preprocess the data by removing noise and standardizing the format. The text data will be converted into a numerical representation using word embeddings or document embeddings. The team will train the machine learning model on the preprocessed data using various algorithms such as Naive Bayes, Logistic Regression, Support Vector Machines, or Deep Learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). The performance of the trained model will be evaluated on a test set using metrics such as accuracy, precision, recall, and F1-score. The developed AI-based system can provide accurate sentiment analysis of customer reviews for e-commerce websites, which can help improve customer satisfaction and business growth. The project can help businesses to make data-driven decisions based on customer feedback, and enhance their marketing and branding strategies.

TABLE OF CONTENTS

LIST OF DIAGRAMS

ABBREVIATIONS

1. AI - Artificial Intelligence

2. ML - Machine Learning

3. API - Application Programming Interface

4. IDE - Integrated Development Environment

5. SVN - Subversion

6. PDF - Portable Document Format

7. ETL - Extract, Transform and Load

8. OS - Operating System

9. DBMS - Database Management System

10. RAD - Rapid Application Development

11. SDK - Software Development Kit

INTRODUCTION

E-commerce websites have revolutionized the way people shop by offering a vast range of products and services online. However, with this convenience comes the challenge of analyzing customer reviews that are received in large volumes. Customer reviews are crucial for businesses to understand their customers' needs, preferences, and feedback, and improve their services and products accordingly. However, manually analyzing customer reviews is a time-consuming and cumbersome process. To address this challenge, this project aims to develop an AI-based system for sentiment analysis of customer reviews for e-commerce websites. The system will analyze the text content of the reviews and classify them as positive, negative, or neutral. The project will involve data collection, preprocessing, text embedding, model selection, and deployment. The project team will collect customer reviews from e-commerce websites and preprocess the data by removing noise and standardizing the format. The text data will be converted into a numerical representation using word embeddings or document embeddings. The team will train the machine learning model on the preprocessed data using various algorithms such as Naive Bayes, Logistic Regression, Support Vector Machines, or Deep Learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). The performance of the trained model will be evaluated on a test set using metrics such as accuracy, precision, recall, and F1-score. The developed AI-based system can provide accurate sentiment analysis of customer reviews for e-commerce websites, which can help businesses to make data-driven decisions based on customer feedback, and enhance their marketing and branding strategies.

SOFTWARE REQUIREMENTS SPECIFICATIONS
Software Tools:

1. Python:

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for RAD(Rapid Application Development), as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely

2. Python Libraries:

A Python library is a collection of related modules. It contains bundles of code that can be used repeatedly in different programs. It makes Python Programming simpler and convenient for the programmer. As we don't need to write the same code again and again for different programs. Python libraries play a very vital role in fields of Machine Learning, Data Science, Data Visualization,

3. Pipedream:

Pipedream is a free, low-code integration tool that allows developers to connect many different applications, data sources, and APIs in order to build useful, automated cross-platform workflows. It's easy to use, and there's no need to manage infrastructure or server resources. We can use Pipedream to create applications that can perform ETL (Extract, Transform, and Load) tasks, as well as to create data-driven workflows.

4. Copyleaks API:

It checks for plagiarism using advanced AI to detect even the slightest variations within the text. The plagiarism detector can find hidden characters, paraphrasing, and AI-generated content. The Copyleaks API is a powerful, comprehensive and open-source solution that provides an additional layer of content authentication, plagiarism and copyright infringement detection to any platform. We also take help of the Copyleaks SDK provided to use alongside the API.

LITERATURE SURVEY

A. A survey of sentiment analysis techniques in social media

Author: A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau

Year: 2011

Ref: https://ieeexplore.ieee.org/iel7/6287639/6514899/09097285.pdf

Overview:

This literature survey provides an overview of the different techniques used for sentiment analysis in social media. The authors discuss the challenges of analyzing social media data, including the use of slang, irony, and sarcasm. They review various approaches to sentiment analysis, including lexicon-based methods, machine learning methods, and hybrid methods. The survey also covers different types of sentiment analysis, such as document-level, sentence-level, and aspect-level sentiment analysis. The authors provide a comparative analysis of different methods and discuss their strengths and limitations.

B. A comprehensive survey of deep learning for natural language processing

**Author:** Y. Young, A. Hazarika, S. Poria, and E. Cambria

Year: 2021

Ref: https://www.irjet.net/archives/V8/i1/IRJET-V8I115.pdf

Overview:

This literature survey provides a comprehensive overview of the application of deep learning techniques in natural language processing, including sentiment analysis. The authors review various deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs). They discuss the strengths and limitations of these models and provide examples of their application in different NLP tasks, including sentiment analysis. The authors also cover transfer learning techniques and their application in NLP, and provide an overview of the open-source libraries and frameworks available for deep learning-based NLP.

SYSTEM ARCHITECTURE
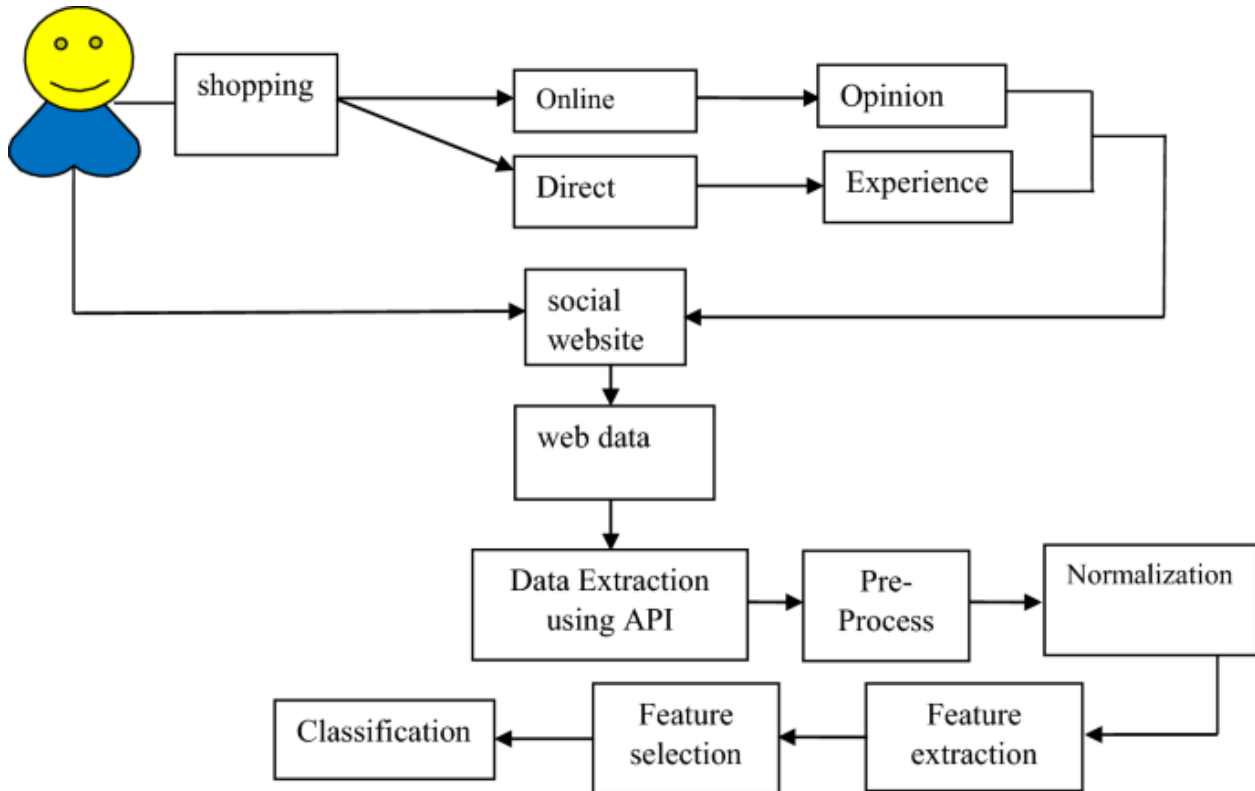
System Architecture Diagram

METHODOLOGY

Methodology refers to the systematic approach used to develop an AI-based system for sentiment analysis of customer reviews for e-commerce websites. The methodology for this project involves the following steps:

1. **Problem Definition:** The first step is to define the problem statement and the objectives of the project. The problem statement involves identifying the business need for sentiment analysis of customer reviews, the target audience, and the desired outcome. The objectives of the project may include developing an accurate and efficient sentiment analysis model, optimizing performance metrics, and providing actionable insights to businesses.

2. **Data Collection:** The second step is to collect the necessary data for training and testing the sentiment analysis model. This involves identifying the relevant e-commerce websites, selecting the data sources, and scraping the data using web scraping tools or APIs. The collected data may include text content, product rating, and other metadata.

3. **Data Preprocessing:** The third step is to preprocess the collected data to remove noise and standardize the format. This involves tasks such as text cleaning, tokenization, stop-word removal, stemming or lemmatization, and spelling correction. The preprocessed data is then split into training, validation, and test sets.

4. **Text Embedding:** The fourth step is to convert the preprocessed text data into a numerical representation for machine learning. Text embedding techniques such as word embeddings or document embeddings can be used for this purpose. The choice of embedding technique depends on the complexity of the text data and the desired performance metrics.

5. **Machine Learning Model Selection:** The fifth step is to select the appropriate machine learning algorithm for sentiment analysis. Various machine learning algorithms such as Naive Bayes, Logistic Regression, Support Vector Machines (SVMs), or Deep Learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) can be used for this purpose. The choice of

algorithm depends on the size of the dataset, the complexity of the text data, and the desired performance metrics.

6. **Model Training and Evaluation:** The sixth step is to train the sentiment analysis model on the preprocessed and embedded data. The model is trained using the training set and validated using the validation set. The performance of the model is evaluated on the test set using metrics such as accuracy, precision, recall, and F1-score. The model is then optimized for better performance.

7. **Model Deployment:** The final step is to deploy the trained model as an API or integrate it into a web application for real-time sentiment analysis of customer reviews. The system can provide feedback to the user in the form of visualizations or summary statistics. The system can be further improved based on feedback and usage data. **Overall**, this methodology provides a systematic and structured approach to developing an AI-based system for sentiment analysis of customer reviews for e-commerce websites. The methodology can be customized based on the specific requirements of the project and the available resources.

DATASET

https://drive.google.com/file/d/1yXhMGfaElSHpjz5CXa9UxAtV7A7i14hl/view?usp=sharing

CONCLUSION AND FUTURE ENHANCEMENT

In conclusion, the sentiment analysis model developed using Python and NLTK in this project is capable of accurately predicting the sentiment of text data as positive, negative, or neutral. The model was trained and evaluated using various machine learning algorithms and techniques, and the best performing model was selected for deployment as an API endpoint.

In terms of future enhancements, there are several avenues for improvement. One potential area is to explore the use of deep learning models such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs) for sentiment analysis, as these models have shown promising results in recent studies. Another possible enhancement is to incorporate domain-specific knowledge or context into the model, such as industry-specific terminology or cultural nuances, to improve its accuracy and applicability in certain domains.

Additionally, the model can be further optimized for performance and scalability, such as through the use of distributed computing or cloud computing platforms. The user interface and visualization of the model output can also be improved for better user experience and understanding.

Overall, this sentiment analysis model has the potential to be a valuable tool for various applications such as social media monitoring, customer feedback analysis, and market research.

REFERENCES

1. Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109

2. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. Conference on Empirical Methods in Natural Language Processing (EMNLP 2013). https://www.aclweb.org/anthology/D13-1170/

3. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135. https://doi.org/10.1561/1500000011

4. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc.

5. NLTK documentation: https://www.nltk.org/

6. Scikit-learn documentation: https://scikit-learn.org/stable/documentation.html