SENTIMENT ANALYSIS AI SYSTEM

A Project Report

submitted in partial fulfillment of the requirements

of

AIML fundamentals with cloud computing with gen AI

by

BALACHANDRAN M

Balachandran.m06@gmail.com

3FE33DECF27047DAF614777D7D6F6714(au922121114018)

Under the Guidance of

Name of Guide (P.Raja, Mater Trainer)

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work. A heartfelt thank you to my supervisor, P.Raja and Mr.Shankar Kannan, whose mentorship extended beyond the academic realm into personal guidance, providing comfort and motivation during tough times. I am also thankful for my friends, who provided both distractions when needed and encouragement when it seemed impossible to continue.

ABSTRACT

Based sentiment analysis involves counting positive and negative words in the text to determine the overall sentiment. Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. The three most popular types, emotion based, fine-grained and aspect-based sentiment analysis (ABSA) all rely on the underlying software's capacity to gauge something called polarity, the overall feeling that is conveyed by a piece of text. AI tools have revolutionized sentiment analysis by automating the process and enhancing its accuracy. By employing machine learning algorithms and NLP techniques, these tools can efficiently process large volumes of text data and extract valuable insights regarding public sentiment. Sentiment analysis is the process of analyzing digital text to determine if the emotional tone of the message is positive, negative, or neutral. Today, companies have large volumes of text data like emails, customer support chat transcripts, social media comments, and reviews. By calculating a sentiment score, companies can better understand the overall sentiment of their audience. One common method in rule

TABLE OF CONTENTS

Chapter 1. Introduction

1.1 Problem Statement

Interpreting the meaning of words in different contexts. AI often struggles with analyzing different sentiments, such as sarcasm or negation, or sentences with multiple sentiments. This makes it challenging to categorize text as positive, negative, or neutral.

1.2 Motivation

Using sentiment analysis in business decision-making can help businesses make informed choices by providing insights into customer preferences, market trends, and brand perception. This allows businesses to tailor their products, services, and marketing strategies to better meet customer needs.

1.3 Objectives

Sentiment analysis is applied to consider customer feedback, employee engagement, survey responses, and product reviews. Monitoring social media, brand reputation management, and customer experience are key areas for successful usage and implementation of sentiment analysis.

1.4 Scope of the Project

Sentiment analysis, or opinion mining, is the computational study of opinions, sentiments, and emotions expressed in text. Its use is growing as it provides valuable insights that can aid in the monetization of products and services.

Chapter 2. Literature Survey

2.1 Previous Work

Previous work in sentiment analysis has largely focused on traditional machine learning algorithms like SVM and Naive Bayes. Recent advancements, however, have shifted towards deep learning techniques, especially models like Long Short-Term Memory (LSTM) and Transformer-based models such as BERT, which better capture contextual nuances.

2.2 Existing Models and Techniques

LSTM and GRU models are effective for sentiment analysis due to their ability to retain long-term dependencies in sequential data. However, they struggle with complex linguistic constructs like sarcasm. SVM performs well in high-dimensional spaces but lacks the flexibility of deep learning models in capturing the subtleties of human language.

2.3 Gaps and Limitations

Despite advances, many models still struggle with complex linguistic features such as sarcasm, irony, and negation. This project aims to address these issues by leveraging state-of-the-art pre-trained models like BERT, fine-tuning them with custom datasets to improve performance on nuanced tasks.

Chapter 3. Proposed Methodology

3.1 System Design

A deep learning-based model utilizing pre-trained BERT will serve as the core sentiment analysis engine. The system will preprocess text data by removing noise, tokenizing it, and passing it through the BERT model for sentiment classification.

3.2 Modules Used

The project involves several modules, including data preprocessing and sentiment classification using AI models. (Remove or modify any sections related to "Face Detection" if not relevant.)

3.3 Data Flow Diagram (DFD)

Data flow diagrams (DFDs) will be used to visualize the data flow from data collection to sentiment classification, illustrating the process aspects of the system.

3.4 Advantages

- **Customer Insights**: Understand customer preferences for better product and service alignment.
- **Data-Driven Decisions**: Inform business strategies with sentiment trends.
- **Real-Time Monitoring**: Quickly respond to feedback and manage brand reputation.
- **Scalability**: Process large datasets efficiently, saving time and resources.
- **Early Issue Detection**: Address problems before they escalate, boosting customer satisfaction.
- **Competitive Edge**: Gain market insights for improved adaptability and customer experience.

3.5 Requirement Specification

- Hardware Requirements
- Software Requirements

Chapter 4. Implementation and Results

4.1 Data Collection and Preprocessing

Collected data from sources such as customer reviews and social media. Preprocessing steps included:

- Data Cleaning: Removed special characters, emojis, and extra whitespace.
- Tokenization: Split text into smaller parts (tokens) for easier processing.
- Normalization and Stop Word Removal: Converted text to lowercase and removed common, non-informative words to focus on key content.
- Lemmatization/Stemming: Reduced words to their base forms for consistency.

4.2 Model Selection and Training

The pre-trained BERT model was chosen due to its ability to capture context. Fine-tuning BERT on a sentiment-specific dataset helped it recognize sentiment nuances. Hyperparameters (like learning rate) were optimized for better accuracy.

4.3 Model Pipeline

The system pipeline processed text through three main steps:

- Input Processing: Tokenization and text formatting.
- Sentiment Analysis: BERT model layers processed the text to identify sentiment.
- Output Classification: Sentiment was categorized as positive, negative, or neutral.

4.4 Evaluation Metrics

Metrics used included:

- Accuracy: Overall correctness of predictions.
- Precision, Recall, and F1-Score: Assessed model reliability, especially for nuanced sentiments.
- Confusion Matrix: Visualized correct vs. incorrect predictions, highlighting challenges with sarcasm and negation.

4.5 Results and Analysis

The model achieved an 85% accuracy. It handled direct sentiment well but had limitations with sarcasm and negation, areas identified for future improvement.

Chapter 5. Discussion and Conclusion

5.1 Key Findings

Key findings indicate that pre-trained models like BERT enhance sentiment analysis accuracy. However, challenges remain, particularly in detecting sarcasm and negation, underscoring the need for model refinement.

5.2 GitHub Link of the Project

https://github.com/Balachandranmurugan/Naan-Mudhalvan-Project.git

5.3 Video Recording of Project Demonstration

https://drive.google.com/file/d/1FiDGsMsrkwtTt3KRFhtUxSKBUHjK6W-V/view?usp=drivesdk

5.4 Limitations

Outline limitations in the model, including potential data-related constraints such as dataset balance and domain variety.

5.5 Future Work

Suggest future improvements, like incorporating multimodal sentiment analysis (text, facial expressions, voice tone) or exploring reinforcement learning to enhance model performance based on user feedback.

5.6 Conclusion

Summarize the overall impact and contributions of the project, emphasizing its application potential in business and customer service contexts.

REFERENCES

Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.