INTERNSHIP PROJECT DOCUMENT

ON "SENTIMENT ANALYSIS OF CUSTOMER REVIEWS"

Submitted by Ganesh Agrahari

Submitted to : **QTechSolutions**

Founder – Kanduri Abhinay

CEO & CTO – Rithin Varma

INTRODUCTION

Sentiment analysis is a key application of Natural Language Processing (NLP) used to determine the sentiment expressed in textual data. This project focuses on analyzing customer reviews of products, categorizing them into positive or negative sentiments. The insights derived from this analysis help businesses understand customer feedback and improve their products and services.

This project follows a supervised learning approach, utilizing NLP techniques and deep learning models to classify sentiment. The final model is deployed as a Streamlit web application.

SOFTWARE DEVELOPMENT LIFE CYCLE (SDLC)

1. Planning

The objective of the project is to develop an AI-based sentiment analysis system that classifies customer reviews into positive or negative categories. The project involves data preprocessing, feature extraction, model training, evaluation, and deployment as an interactive web app.

2. Defining Requirements

Software & Hardware Requirements:

- **Operating System:** Windows/Linux/MacOS
- **Programming Language:** Python
- **Libraries Used:** Pandas, NumPy, NLTK, Transformers (Hugging Face), Scikit-learn, Streamlit,tqdm
- **Development Environment:** Jupyter Notebook, VS Code
- **Hardware Requirements:** Recommended 8GB RAM, Intel i5+ Processor

3. Designing

The system design includes the following key components:

- **Data Preprocessing:** Removing noise, tokenization, stopword removal, and lemmatization.

- **Feature Extraction:** Converting text into numerical representation using Transformers-based embeddings.
- **Model Selection:** Fine-tuning Roberta, a pre-trained Transformer model.
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-score.
- **Deployment:** Developing an interactive Streamlit web application for real-time sentiment analysis.

---

4. Building (Implementation)

The implementation follows a structured machine-learning workflow:

- **Load Dataset:** Extract Amazon customer reviews from a CSV file.
- **Text Processing:** Clean and preprocess the text data.
- **Feature Engineering:** Convert text into embeddings using the Hugging Face Transformers pipeline.
- **Train Model:** Fine-tune Roberta for sentiment classification.
- **Evaluate Model:** Use confusion matrix, accuracy score, and F1-score for performance assessment.
- **Deploy Using Streamlit:** Develop a user-friendly web interface for sentiment analysis.

---

5. Testing

The sentiment analysis model was tested using different evaluation techniques:

- **Unit Testing:** Verified text preprocessing, feature extraction, and model performance.
- **Performance Testing:** Evaluated the model's ability to process large datasets efficiently.
- **Edge Cases:** Ensured robustness against misspellings, slang, and ambiguous sentiments.

---

6. Deployment

- The final trained model was deployed as a **Streamlit Web App**.
- Users can input text and receive sentiment predictions with confidence scores.
- The app features a **dark-themed UI** and real-time sentiment visualization.

PROCEDURE AND METHODS USED

1. **Data Preprocessing:** Cleaning the dataset, removing stopwords, and applying tokenization.
2. **Feature Extraction:** Utilizing Transformer-based embeddings from Roberta.
3. **Model Training:** Fine-tuning Roberta for sentiment classification.
4. **Model Evaluation:** Assessing accuracy, precision, recall, and F1-score.
5. **Deployment:** Developing an interactive Streamlit app with sentiment visualization.

ALGORITHM (Roberta - TRANSFORMERS)

1. Load the dataset.
2. Preprocess the text data (remove stopwords, tokenization, lemmatization).
3. Convert text into Transformer embeddings using the Hugging Face pipeline.
4. Fine-tune Roberta on the dataset.
5. Evaluate the model using a confusion matrix and F1-score.
6. Develop a Streamlit web application for real-time analysis.
7. Deploy and test the application.

FUTURE SCOPE

1. **Advanced Deep Learning Models:** Implementing LSTMs and GPT-based models for better accuracy.
2. **Multilingual Sentiment Analysis:** Expanding support for multiple languages.
3. **Real-Time Analysis:** Enhancing performance for live customer feedback.
4. **Feature Expansion:** Incorporating additional metadata like review length, rating, and customer history.
5. **Cloud Deployment:** Deploying the model on AWS/GCP for scalability.

ADVANTAGES

- Provides automated insights from customer reviews.
- Enhances customer experience by identifying common issues.
- Helps businesses improve their products based on sentiment trends.
- Scalable and efficient for analyzing large datasets.
- Interactive web-based visualization improves usability.

REFERENCES

1. Sentiment Analysis - NLTK & Hugging Face Documentation
2. Fine-tuning Roberta - Hugging Face Official Guide
3. Streamlit for Web Deployment - Documentation
4. Feature Extraction for NLP - Research Papers & Articles

CONCLUSION

This project successfully applies deep learning techniques using **Roberta** for sentiment analysis of reviews. The final model is deployed as a **Streamlit Web App**, enabling real-time sentiment predictions with a confidence score. Future enhancements include expanding the model's capabilities with advanced

Transformer-based architectures and deploying it on cloud platforms for greater scalability.

This project highlights the power of **AI in understanding customer sentiments** and improving business decision-making through automated text analysis.