Project Report

AI-Driven Sentiment Analysis using NLP and ANN

1. Introduction

Sentiment analysis is a key application of Natural Language Processing (NLP) that focuses on identifying opinions or emotions expressed in textual data. With the rapid growth of online platforms, customer reviews have become an important source of feedback for businesses. This project aims to analyze restaurant reviews and automatically classify them as **positive** or **negative** using NLP techniques and an Artificial Neural Network (ANN).

The project demonstrates a complete machine learning workflow, starting from raw text preprocessing to model training and prediction on new, unseen reviews.

---

2. Problem Statement

Restaurant owners and customers generate a large volume of textual reviews, making manual analysis impractical. The objective of this project is to build an automated system that can:

- Understand textual restaurant reviews

- Convert unstructured text into numerical features

- Predict the sentiment polarity (positive or negative) accurately

---

3. Dataset Description

The dataset used in this project consists of **1000 restaurant reviews**. It contains two columns:

- **Review**: Textual feedback provided by customers

- **Liked**: Target label where 1 represents a positive review and 0 represents a negative review

This is a **binary classification problem**.

---

4. Methodology

The project follows a structured pipeline consisting of NLP preprocessing, feature extraction, and model training.

4.1 Text Preprocessing (NLP)

Natural Language Processing techniques are applied to clean and normalize the raw text data:

- Removal of special characters and punctuation using regular expressions

- Conversion of text to lowercase

- Tokenization of sentences into individual words

- Removal of stopwords to reduce noise

- Stemming using Porter Stemmer to reduce words to their root form

These steps help in improving model performance by reducing irrelevant information.

4.2 Feature Extraction

After preprocessing, the textual data is transformed into numerical form using the **Bag of Words (BoW)** model implemented through **CountVectorizer**. A fixed vocabulary size is selected to represent the most frequent and relevant words. This step converts each review into a numerical feature vector that can be processed by the neural network.

4.3 Model Building – Artificial Neural Network

An Artificial Neural Network (ANN) is used for sentiment classification. The network consists of:

- An input layer corresponding to the Bag of Words features

- Hidden layers with ReLU activation functions

- An output layer with a sigmoid activation function for binary classification

The model is trained using the **Adam optimizer** and **binary cross-entropy loss function**.

---

5. Model Training and Evaluation

The dataset is split into training and testing sets to evaluate model performance. The ANN is trained over multiple epochs to learn patterns associated with positive and negative sentiments. Model performance is evaluated using accuracy and prediction results on test data.

---

6. Model Persistence and Prediction

After training, the ANN model and the CountVectorizer are saved to disk. This allows the system to reuse the trained components without retraining. During prediction, new reviews undergo the same preprocessing and feature extraction steps before being passed to the trained model for sentiment prediction.

---

7. Results and Observations

The trained ANN model is able to effectively classify restaurant reviews into positive and negative categories. The use of classical NLP techniques combined with ANN provides a reliable baseline approach for sentiment analysis tasks.

---

8. Applications

- Customer feedback analysis for restaurants

- Opinion mining from online review platforms

- Market research and customer satisfaction analysis

---

9. Limitations

- Uses Bag of Words, which ignores word order and context

- Does not handle sarcasm or complex language patterns

- Limited to binary sentiment classification

---

10. Future Enhancements

- Use TF-IDF or word embeddings for better feature representation

- Apply deep learning models such as LSTM or Transformers

- Extend the system to multi-class sentiment analysis

- Deploy the model as a web or mobile application

---

11. Conclusion

This project successfully demonstrates an AI-driven sentiment analysis system using NLP and ANN. By combining text preprocessing, feature extraction, and neural network-based classification, the system effectively predicts sentiment from restaurant reviews. The project highlights the practical application of NLP and machine learning techniques in real-world text analysis problems.