ReviewRadar

Enhancing Product Review Analysis with Advanced NLP Techniques

CORPORATE GURUKUL

Index

CORPORATE GURUKUL

Team Introduction

1. Sabeeh Israr
2. Ashwin Singh
3. Godavari Venkata Sai Tharun
4. Kiran Shakti
5. Saivikraman

CORPORATE GURUKUL

Problem Statement

E-commerce platforms rely heavily on customer reviews to inform potential buyers about product quality and performance. However, the presence of fake reviews can mislead customers and undermine trust in the platform. Additionally, understanding the sentiment of genuine reviews can provide valuable insights into customer satisfaction and product performance. This project aims to develop a robust system that can:

1. **Perform Sentiment Analysis**: Classify the sentiment of product reviews (positive, negative) to provide an overall sentiment score for products.
2. **Detect Fake Reviews**: Identify and filter out fake reviews using advanced machine learning techniques to ensure that only genuine customer feedback is considered.

CORPORATE GURUKUL

Abstract

- **Importance of Online Reviews**: Online reviews are essential for consumer decisions but are often compromised by fake reviews
- **Project Objective**: Develop a system to analyze reviews and understand customer sentiment
- **Techniques Used**: Use NLP techniques and machine learning methods to determine overall sentiment and detect fake reviews by analysing existing reviews
- **Business Benefits**: Enhance customer trust and support informed purchasing decisions by presenting accurate product information. Enhance information available to companies about product perception
- **Applications**: Can be used by e-commerce platforms, review aggregators, and businesses relying on online feedback for product assessment

Business Use-Cases

E-Commerce Platforms:

- **Authenticity Verification**: Identify and remove fake reviews to enhance trust and improve shopping experiences
- **Product Recommendations**: Use genuine sentiment data to offer accurate product recommendations, boosting sales

Brand Reputation Management:

- **Reputation Monitoring**: Gain insights into customer perceptions and address issues raised in genuine reviews
- **Competitive Analysis**: Analyze competitors' review data for strategic decision-making and competitive positioning

Market Research Firms:

- **Consumer Insights**: Extract trends and insights from review data to inform product development and marketing strategies
- **Campaign Evaluation**: Assess marketing campaign effectiveness by comparing sentiment data before and after campaigns

Customer Service:

- **Issue Resolution**: Quickly identify and resolve common issues highlighted in reviews to improve customer satisfaction
- **Personalized Support**: Offer personalized support based on individual customer sentiments

Models Used

1. **Logistic Regression**

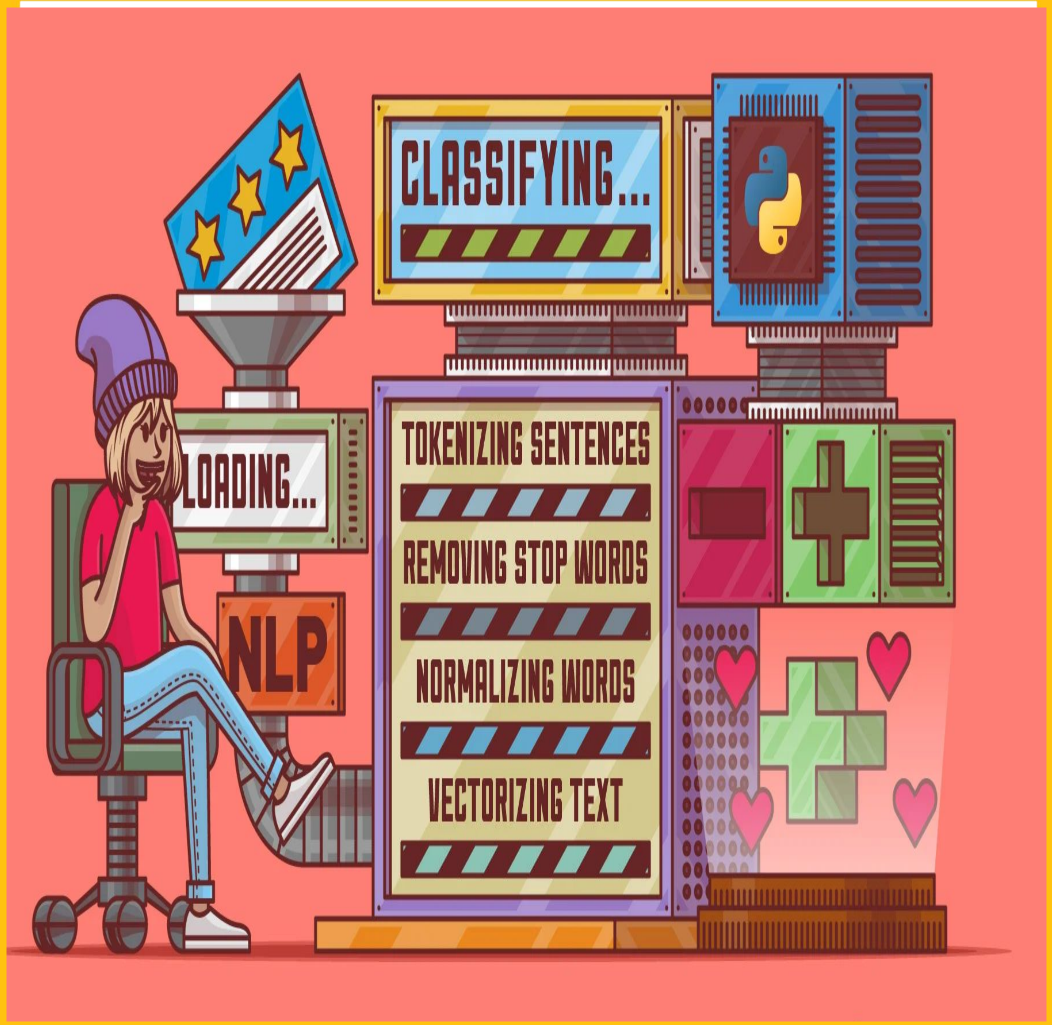
Logistic regression is a statistical method well-suited for tasks like sentiment analysis where you want to categorize text data (reviews) into different classes (positive & negative)

2. **Support Vector Machine (SVM)**

SVM is a classification algorithm that excels at separating data points into distinct categories. In our case, it learns to differentiate between real and fake reviews based on extracted features

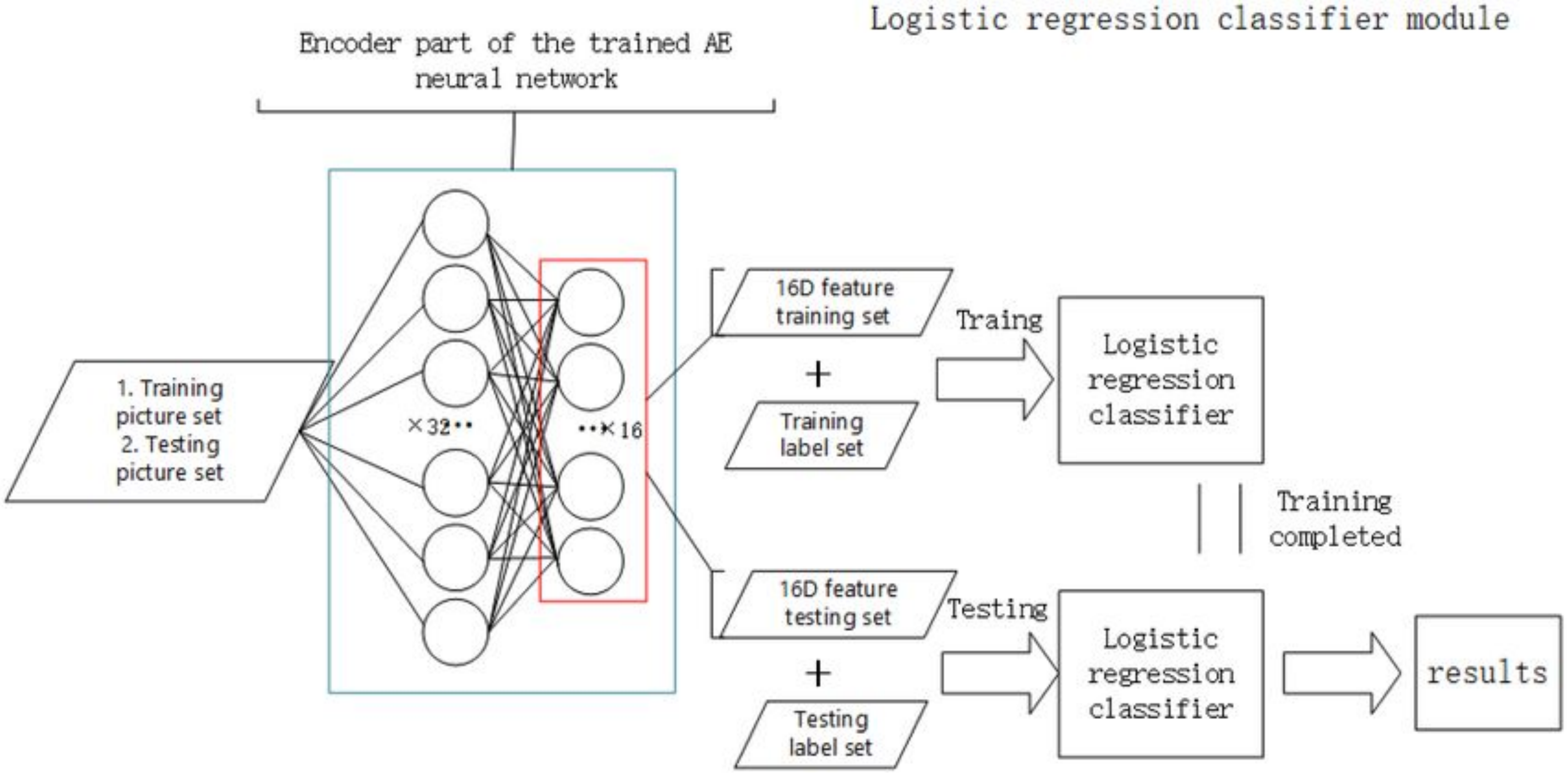
3. **Long Short Term Memory (LSTM)**

LSTM networks are a specialized type of Recurrent Neural Network (RNN). Using it for sentiment analysis leverages its ability to capture long-term dependencies in textual data, making it highly effective in understanding context and nuances in language. This capability results in more robust and accurate sentiment predictions compared to traditional models



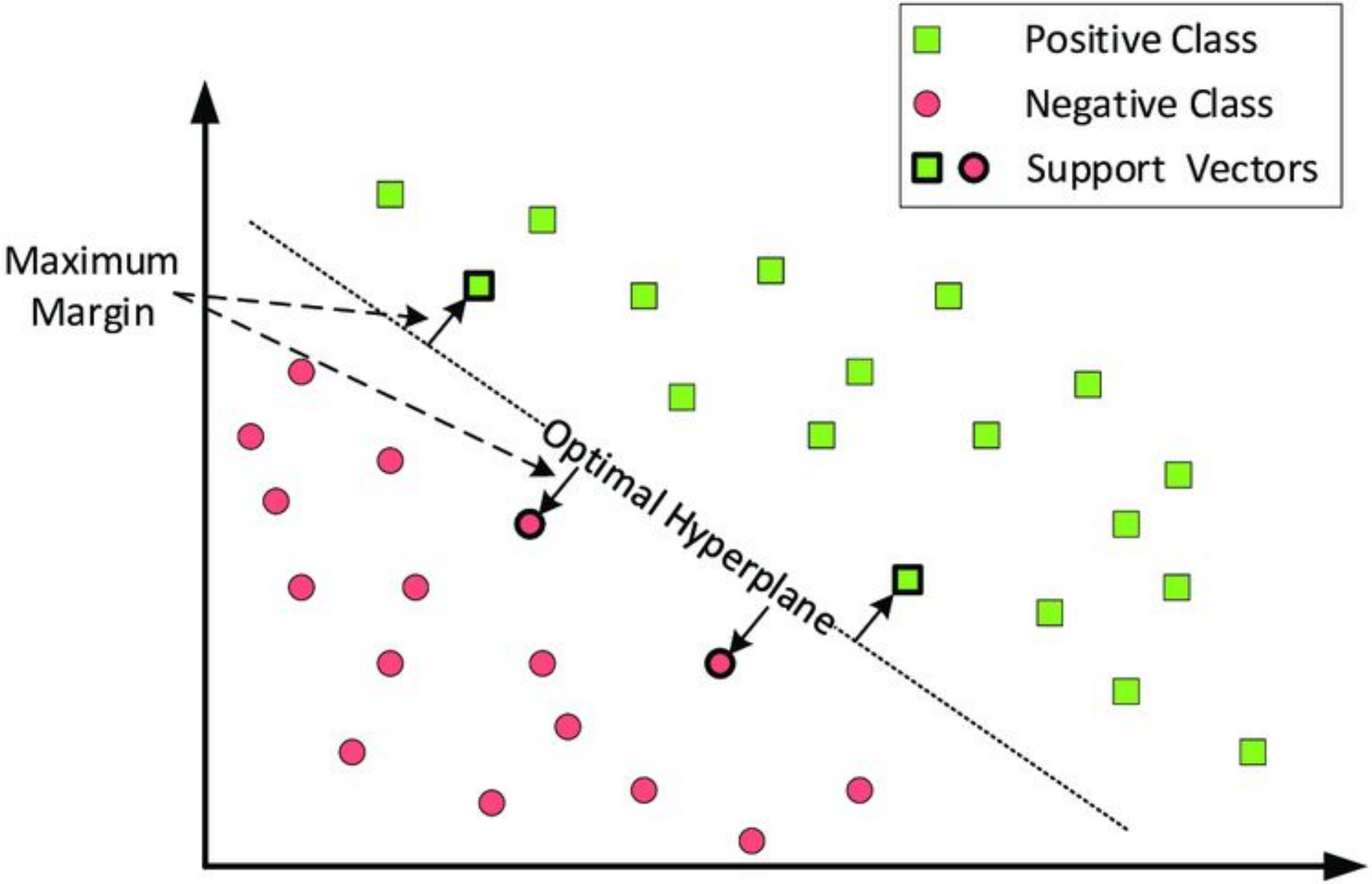
CORPORATE
GURUKUL

Logistic Regression



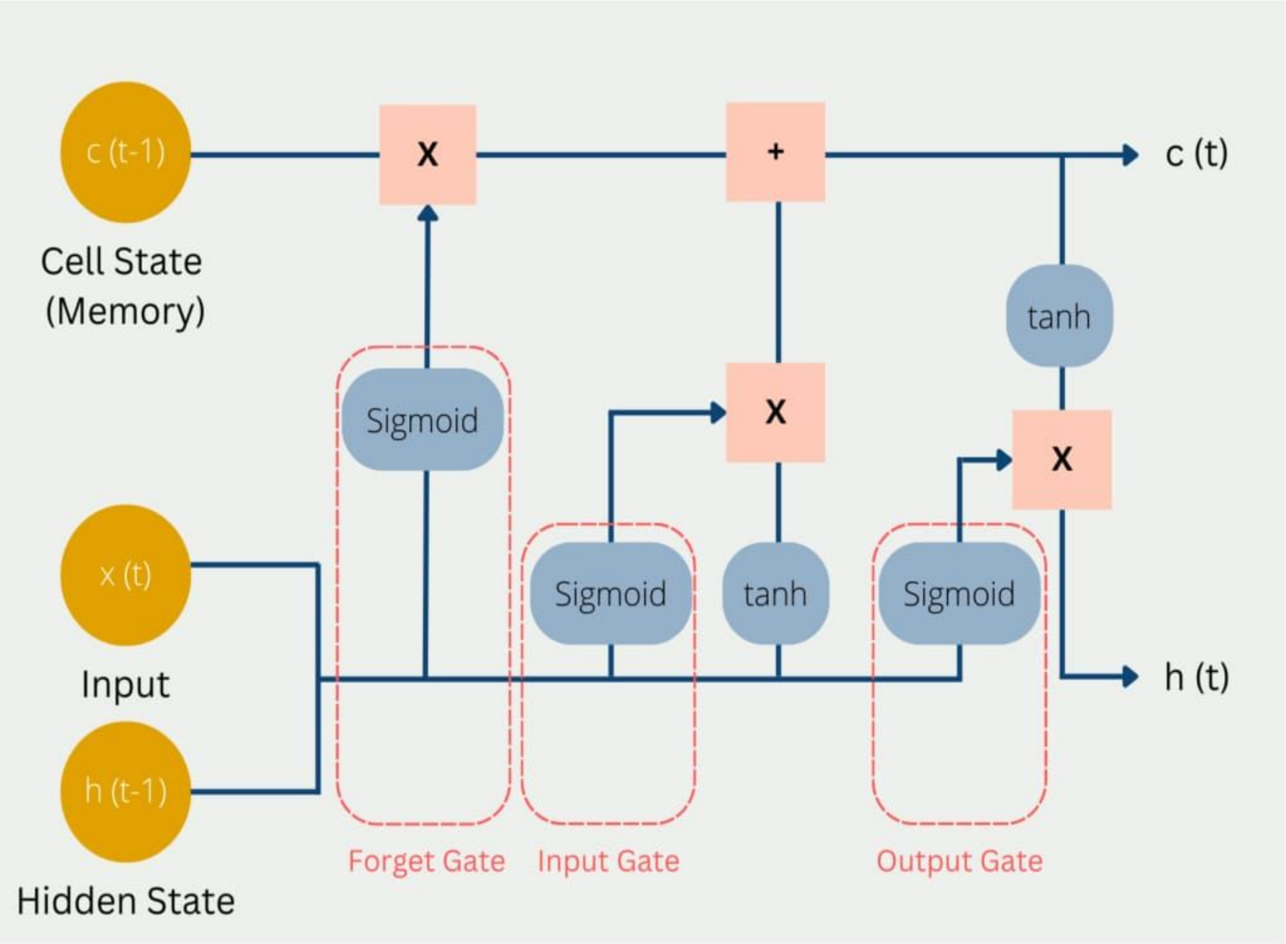
CORPORATE GURUKUL

Support Vector Machine



CORPORATE
GURUKUL

Long Short Term Memory



CORPORATE
GURUKUL

Model Comparison - Sentiment Analysis

1. Logistic Regression

Logistic Regression	
Training Accuracy (0.8)	0.93
Testing Accuracy (0.2)	0.90

Sentiment/ Metric	Precision	Recall	F1 - Score	Support
Bad	0.92	0.95	0.93	6646
Good	0.88	0.81	0.84	2953

2. Recurrent Neural Network (Long Short Term Memory - LSTM)

LSTM	
Training Accuracy (0.8)	0.90
Testing Accuracy (0.2)	0.89

Sentiment/ Metric	Precision	Recall	F1 - Score	Support
Negative	0.94	0.89	0.92	6644
Positive	0.79	0.87	0.83	2967

CORPORATE GURUKUL

Model Comparison - Fake Review Detection

1. Support Vector Machine (SVM)

Support Vector Machine	
Training Accuracy (0.8)	0.92
Testing Accuracy (0.2)	0.87

Label/ Metric	Precision	Recall	F1 - Score	Support
-1 (real)	0.88	0.86	0.87	4006
1 (fake)	0.87	0.88	0.87	4081

2. Recurrent Neural Network (Long Short Term Memory - LSTM)

LSTM	
Training Accuracy (0.8)	0.98
Testing Accuracy (0.2)	0.96

Label/ Metric	Precision	Recall	F1 - Score	Support
-1 (fake)	0.96	0.97	0.96	2013
+1 (real)	0.97	0.96	0.96	2031

Data Collection and Processing

1) **Fake Review Detection with SVM**

Dataset with 40K reviews - 20K fake and 20K real ones

2) **Sentiment Analysis**

A Kaggle dataset with approximately 53K Amazon reviews along with product rating

Word embeddings:

- Bag of Words for Logistic Regression and SVM
- GloVe and Word2Vec for LSTM

Data processing involved cleaning, tokenization, stopword removal and lemmatization

CORPORATE GURUKUL

Challenges Faced

1) Fake Review Detection

Limited Data Availability: limited corpus of training data restricts our model's ability to correctly identify fake reviews

Increased capabilities of LLMs: Computer generated reviews may become increasingly harder to distinguish from genuine human reviews

2) Sentiment Analysis

Sarcasm and Irony: Detecting sentiment in reviews containing sarcasm or irony is a challenge

Data Quality: Noisy and imbalanced datasets can lead to poor classification performance

CORPORATE GURUKUL

Our Learnings

Data Preprocessing:

- Understood techniques for cleaning and preparing review datasets for analysis
- Learned methods for handling text data - cleaning, tokenization, lemmatization, and word embedding

Logistic Regression:

- Gained knowledge of logistic regression for binary classification in sentiment analysis
- Learned how to implement and evaluate logistic regression models using accuracy, precision, recall, and F1 score metrics

Support Vector Machines (SVMs):

- Understood the principles of SVMs and their application in fake review detection
- Learned to implement SVM models for classification tasks and optimize their performance through hyperparameter tuning

Long Short Term Memory Architecture (LSTM):

- Understood the working of LSTM networks and their application to analyze text data
- Learned to implement LSTM models for text classification tasks

Future Enhancements

1. Expand data sources and address imbalance in datasets
2. Review Summarization

CORPORATE GURUKUL

Conclusion

This project has made significant progress in developing a system for sentiment analysis and fake review detection of product reviews.

For sentiment analysis, the first model built with logistic regression demonstrates promising results with high accuracy (around 90%) on both training and testing data, suggesting good generalizability. For our second model, we used LSTM and found that the training and testing performance of both models is comparable.

The fake review detection model, built with support vector machines, achieved a high training and testing accuracy (87%) as well, whereas LSTM performs much better with a testing accuracy of 96%.

For all four models, the classification report demonstrates that their accuracy is robust and that they are effective at their respective classification functions as measured by metrics like precision, recall and F1 score.

CORPORATE
GURUKUL

Thank You

CORPORATE
GURUKUL