

Reviews Sentiment Analysis using Text Classification

MSc in AI - Machine Learning Assignment

February 19, 2025

Introduction

Sentiment analysis is a fundamental task in Natural Language Processing (NLP), where text data is classified based on its emotional tone. In this study, we investigate cross-domain sentiment classification by training a model on one dataset (IMDb or Amazon reviews) and testing it on another. The goal is to evaluate how well sentiment knowledge transfers across different domains.

Dataset and Preprocessing

The datasets used in this study include:

- IMDb movie reviews: A collection of labeled movie reviews.
- Amazon reviews: A collection of product reviews labeled as positive or negative.

Data preprocessing included:

- Text cleaning (removal of punctuation, stopwords, and special characters)
- Tokenization and vectorization
- Feature extraction using TF-IDF and word embeddings

Model Training and Testing

Two experimental setups were designed:

1. Training on IMDb reviews and testing on Amazon reviews.
2. Training on Amazon reviews and testing on IMDb reviews.

A variety of machine learning classifiers were employed, including:

- Logistic Regression
- Support Vector Machines (SVM)
- Random Forest
- Naive Bayes

Results

Confusion Matrices

Confusion matrices provide insights into the performance of classification models by showing the distribution of predicted versus actual labels.

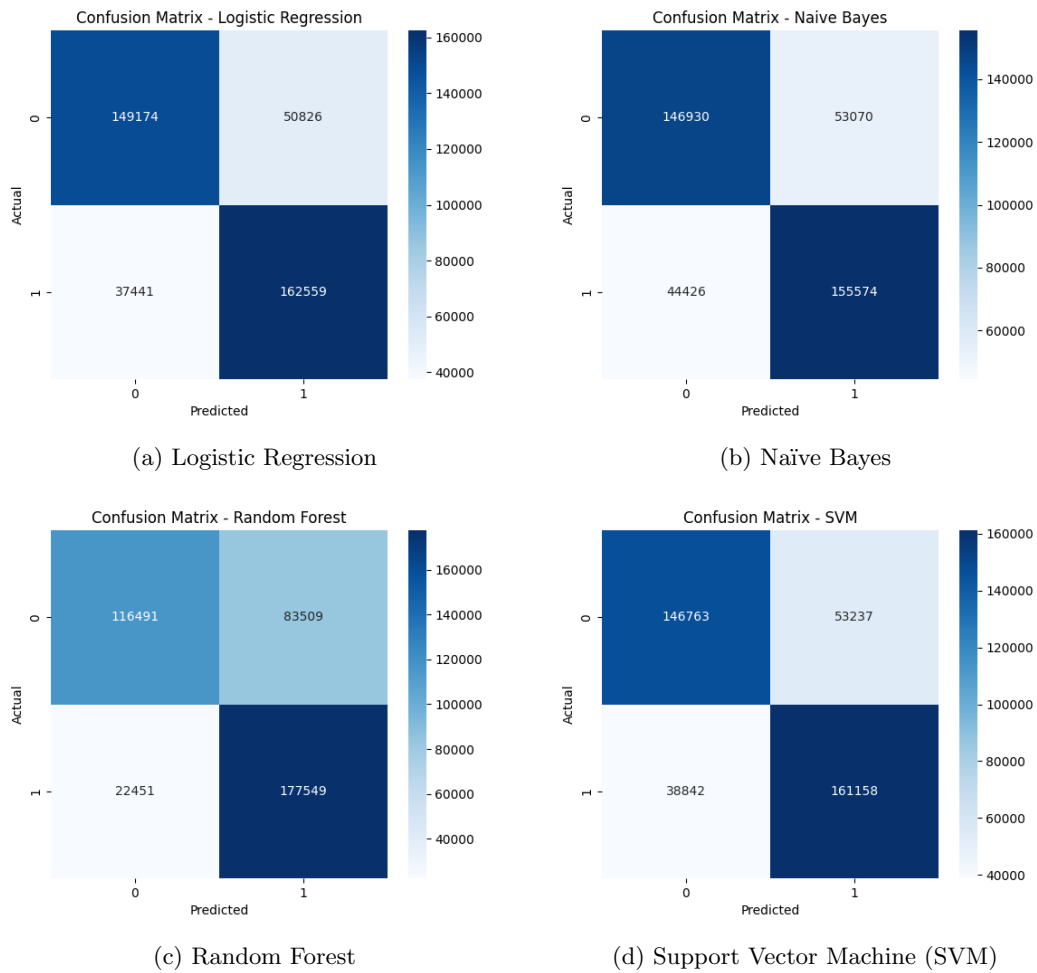


Figure 1: Confusion Matrices for different classifiers

Precision-Recall Curves

Precision-Recall curves help evaluate model performance, especially in imbalanced datasets.

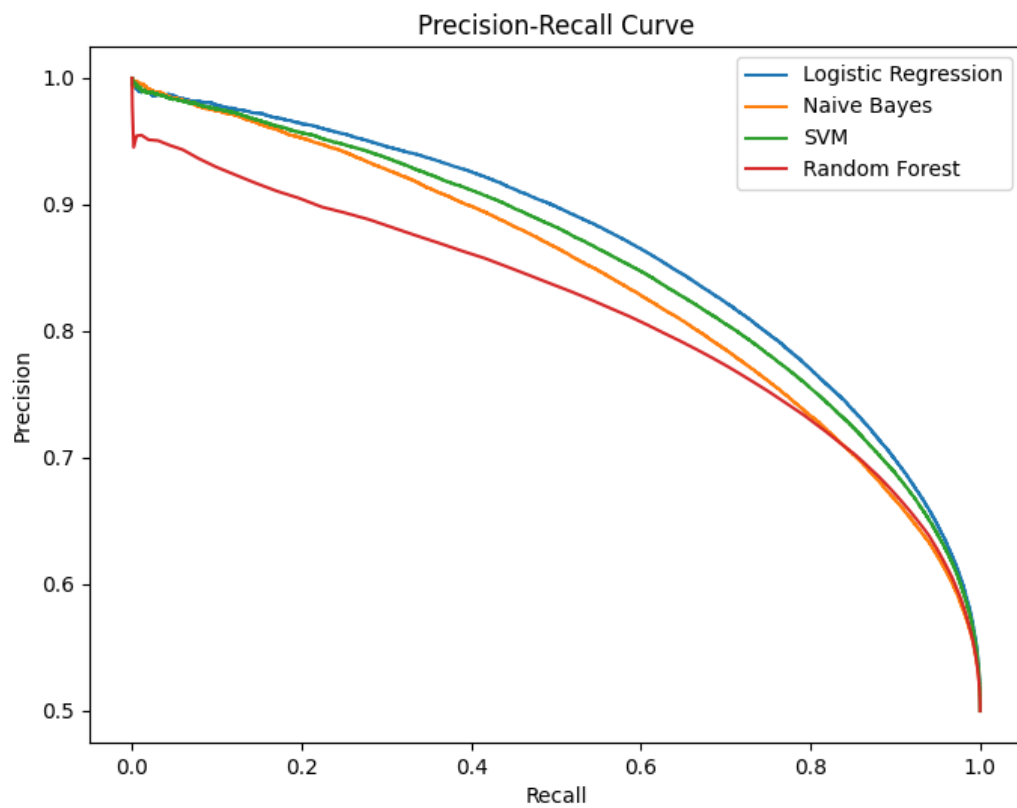


Figure 2: Precision-Recall Curves for sentiment classification.

ROC Curves

The ROC (Receiver Operating Characteristic) curve illustrates the trade-off between true positive rate and false positive rate.

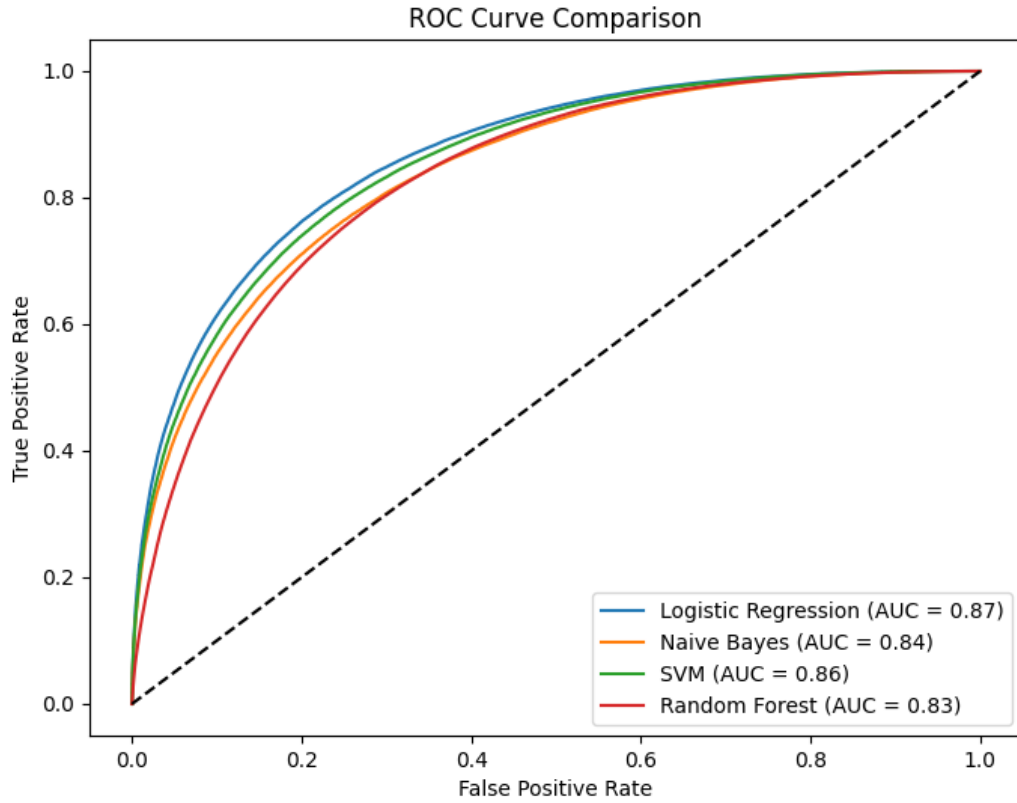


Figure 3: ROC Curves for sentiment classification.

Performance

Performance was evaluated using accuracy, precision, recall, and F1-score. The results are summarized in Table 1.

| Model | F1-score(Class 0) | F2-score (Class 1) | Accuracy | Recall(Class 0) | Recall(Class 1) |
|---------------------|-------------------|--------------------|----------|-----------------|-----------------|
| Logistic Regression | 0.77 | 0.79 | 0.78 | 0.74 | 0.82 |
| Naive Bayes | 0.75 | 0.76 | 0.76 | 0.73 | 0.78 |
| SVM | 0.76 | 0.78 | 0.77 | 0.73 | 0.80 |
| Random Forest | 0.68 | 0.77 | 0.73 | 0.57 | 0.89 |

Table 1: Cross-domain sentiment analysis results.

Performance was evaluated using accuracy across the models. The results are summarized in Table 2.

| Model | Train on IMDB, Test on Amazon | Train on Amazon, Test on IMDB |
|---------------------|-------------------------------|-------------------------------|
| Logistic Regression | 0.78 | 0.54 |
| Naive Bayes | 0.76 | 0.56 |

Table 2: Cross-domain sentiment analysis results.

In Figure 4 are presented the top 20 most important words across models.

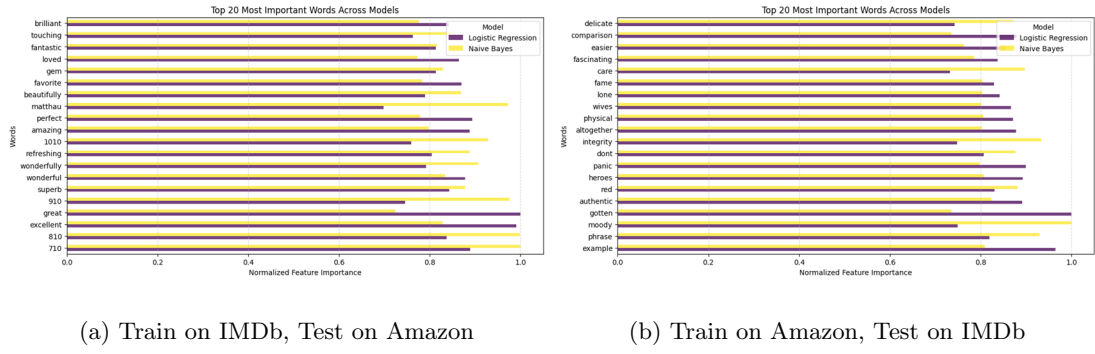


Figure 4: ROC Curves for sentiment classification.