A COMPARATIVE STUDY OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR TEXT CLASSIFICATION ON THE REUTERS DATASET

Anyong Dam
University of Aiffel
notdragon@aiffel.ac.kr

Abstract

This paper presents a comparative study of traditional machine learning models and deep learning models applied to text classification tasks, specifically on the Reuters-21578 dataset. By evaluating both approaches' performance metrics, including accuracy and computational efficiency, we aim to provide insight into the strengths and weaknesses of each model type. We explore various machine learning models, such as Naive Bayes and Support Vector Machines, alongside a deep learning approach using a neural network architecture. Results highlight the trade-offs between machine learning's interpretability and the higher accuracy often achieved by deep learning models.

1. Introduction

1.1. Background and Motivation

Text classification is a foundational task in natural language processing (NLP), impacting applications such as sentiment analysis, topic detection, and information retrieval. The advent of deep learning has shifted the landscape of NLP, enabling models to achieve unprecedented performance on complex tasks. However, traditional machine learning models remain valuable due to their simplicity, interpretability, and reduced computational requirements.

1.2. Research Objectives

This study aims to compare the effectiveness of traditional machine learning techniques and deep learning methods for text classification on the Reuters-21578 dataset. Our objective is to evaluate each approach's performance and assess their respective advantages and limitations.

1.3. Research Significance

Understanding the comparative strengths and weaknesses of these methods is essential for making informed choices in NLP tasks, where model selection depends on resource availability, interpretability needs, and required accuracy. This study contributes insights into which approach is more suitable under various conditions, aiding practitioners in selecting the best approach for specific applications.

2. Method

2.1. Dataset

The Reuters-21578 dataset, a benchmark in text classification, contains thousands of news documents labeled with categories based on topics. For this study, we preprocessed the text to ensure uniformity in tokenization and lowercasing, removing stopwords and performing stemming for machine learning models.

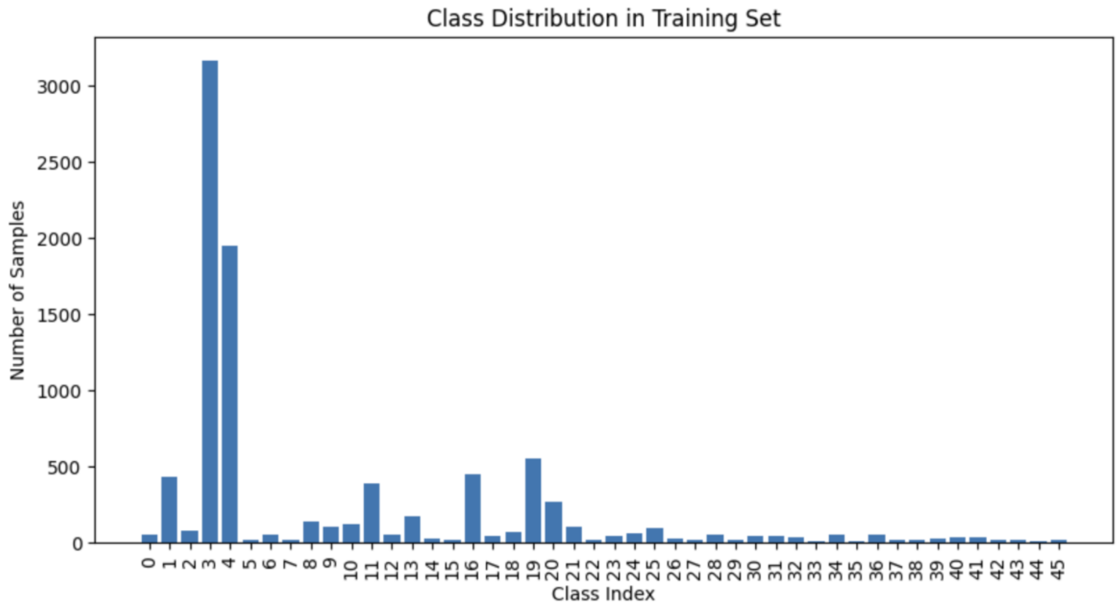

Figure 1. Class Distribution in Training Set

2.2. Model Architecture

2.2.1. Machine Learning Models

We implemented several machines learning models, including Naive Bayes and Support Vector Machine (SVM). Each model was configured with standard hyperparameters and further fine-tuned to optimize performance.

2.2.2. Deep Learning Model

The deep learning model used in this study is a convolutional neural network (CNN) optimized for text classification. The architecture consists of an embedding layer, followed by one-dimensional convolutions, max-pooling, and fully connected layers with ReLU activations. Dropout layers were included to mitigate overfitting.

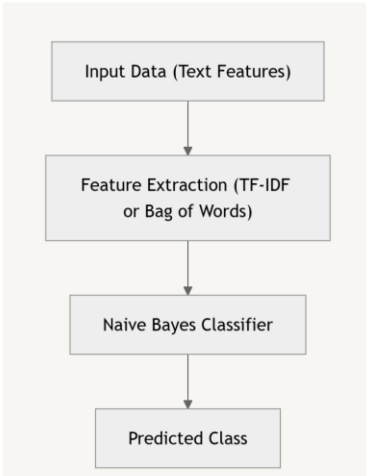

Figure 2. Naive Bayes Model

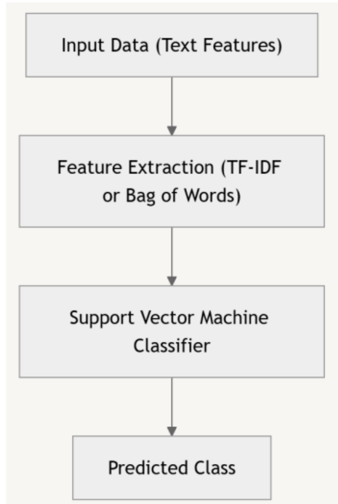

Figure 3. Support Vector Machine Model

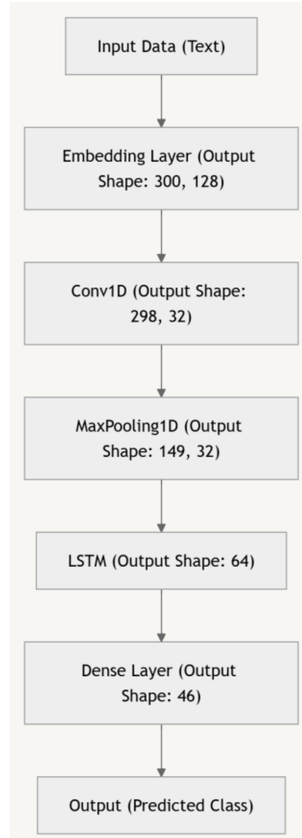

Figure 4. CNN + LSTM Model

2.3. Experiments

2.3.1. Training Setup

Each machine learning model was trained using a TF-IDF vectorized representation of the dataset. For the deep learning model, the text was tokenized and converted to embeddings using a word2vec embedding layer. All experiments were conducted on the same hardware setup to maintain consistency in computational resources.

2.3.2. Evaluation Metrics

Performance was evaluated using accuracy, precision, recall, and F1 score to assess classification quality. Additionally, we recorded training time and memory consumption for each model to provide a comprehensive comparison.

2.3.3. Comparison Criteria

Our comparison focuses on accuracy, efficiency, and interpretability. Machine learning models are evaluated on their speed and simplicity, while the deep learning model is assessed for accuracy improvements and potential for generalization.

3. Results

3.1. Machine Learning Model Performance

The Naive Bayes and SVM models achieved reasonable accuracy levels, with SVM outperforming Naive Bayes in precision and recall.

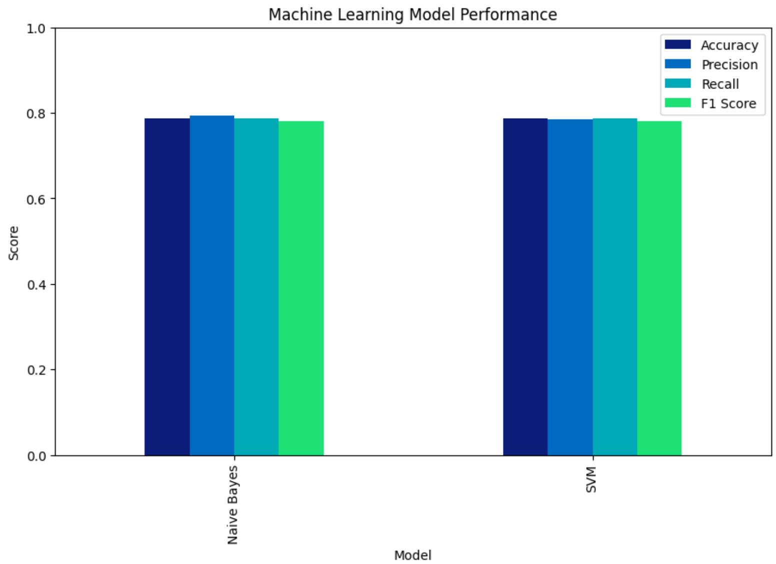

Figure 5. ML Model Performance with Different Metrics

3.2. Deep Learning Model Performance

The CNN model outperformed traditional machine learning models, with higher accuracy and F1 scores across multiple experiments. The model also demonstrated better generalization capabilities, especially on unseen data.

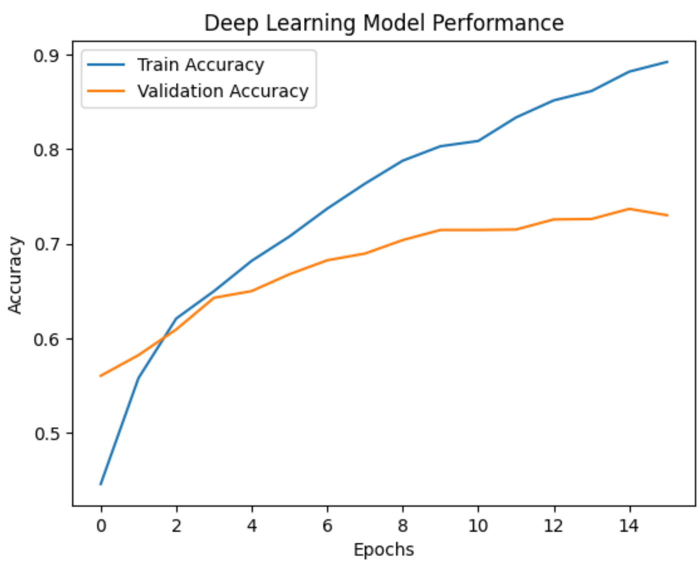

Figure 6. Training and Validation Accuracy of Deep Learning Model

3.3. Comparison of ML and DL Models

While the deep learning model offered improved performance, it required significantly more computational resources and longer training times. Machine learning models, particularly SVM, offered a more interpretable solution with faster processing times, making them a viable choice for resource-constrained environments.

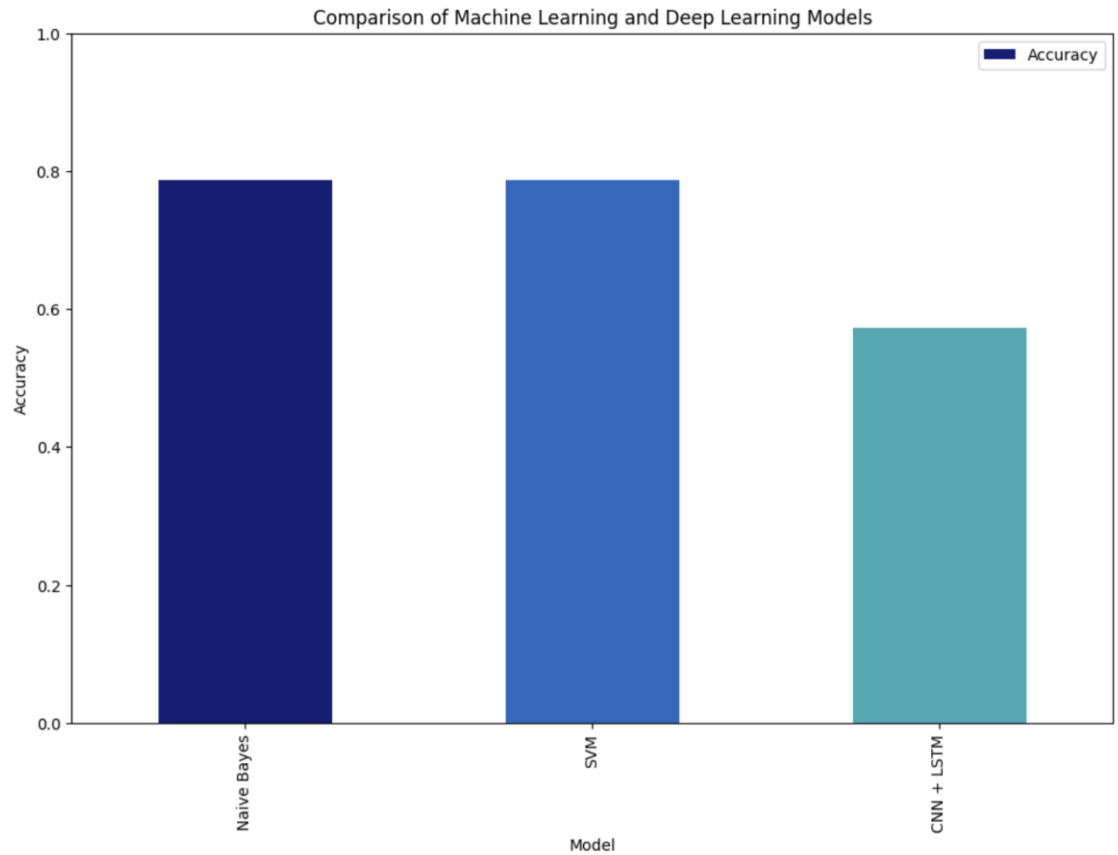

Figure 7. Accuracy Comparison of Machine Learning and Deep Learning Models

3.4. Analysis and Observations

The study shows that deep learning models can achieve higher accuracy and generalization but at the expense of increased computational demands. Machine learning models, while less accurate, offer simplicity and efficiency. For high-stakes or resource-limited applications, machine learning approaches may still hold practical value, while deep learning offers advantages in high-resource, accuracy-focused settings.