Comparative Analysis of Machine Learning and Deep Learning Algorithms for Sentiment Polarity Detection.

-Project By

Raveesh R- 21011101099- 21110282 - 3rd year AI&DS-B

Samyuktha S- 21011101106 -21110237 - 3rd year AI&DS-B

**Introduction:**

Sentiment analysis, or opinion mining, holds significant importance in understanding public sentiment and customer feedback. This research paper concentrates on the task of sentiment polarity detection and conducts a comparative study between conventional machine learning algorithms and cutting-edge deep learning methods. The study leverages the Amazon reviews dataset as its foundational data source. Following text data preprocessing, a range of machine learning techniques and hybrid models are applied to categorize sentiment. Notably, the study highlights the superiority of CNN, achieving an 85% accuracy rate. As part of future research, the paper suggests delving into more advanced deep learning techniques like GRU, Transformers and BERT Model to further enhance predictive accuracy.

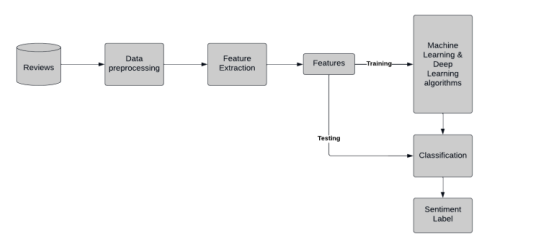
**Dataset:**

For this research paper the experimental dataset is the amazon reviews dataset, which includes the reviews from Amazon selected from Hugging Face. The dataset contains 36 million reviews out of which 18 million are positive comments and the remaining are negative comments. We will be using a subset of the dataset for our research purposes to reduce the computational complexity.

Link: <https://huggingface.co/datasets/amazon_polarity>

**Dataset Sample:**



**Proposed Workflow:**

**Base model and architecture:**

This is the base model architecture taken from the research paper containing sets of Conv2D, BatchNormalisation and MaxPooling layers followed by a flattening layer.

Accuracy of the base model: 85%

r

**Results of the base model:**

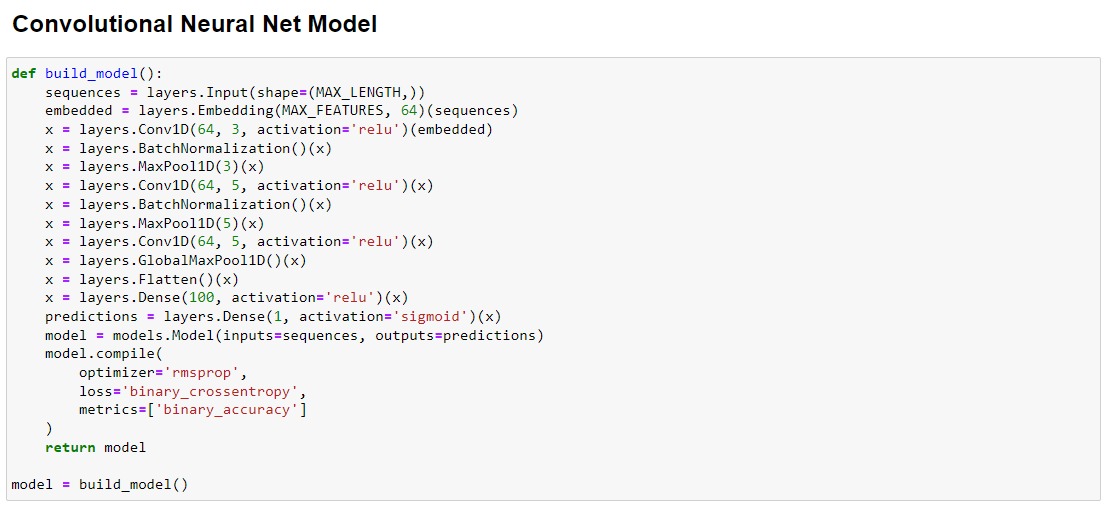
The study compared machine learning and deep learning models with SVM and CNN achieving the highest accuracy, respectively. Naive Bayes outperformed Random Forest, and hybrid models performed better than BiLSTM. Deep learning showed superior learning capacity, with CNN being the best-performing model overall, with an accuracy of 85%.

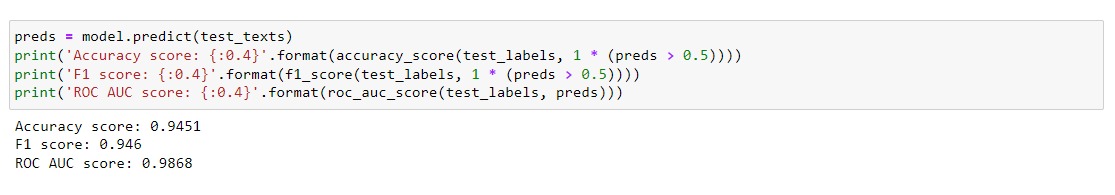
**Refinned model:**

We have implemented an improved model using CNN and RNN that has yielded an accuracy of 95%. The architecture of our model is presented below.

The model can be accessed using the link: https://github.com/raveeshrajkumar/Amazon\_Sentiment\_Analysis

CNN architecture used:

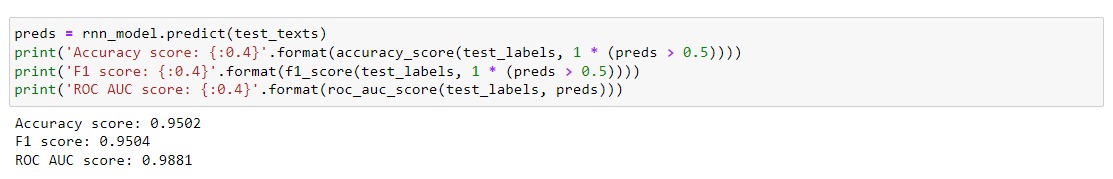




The CNN model has yielded an accuracy of 94.5%.

RNN architecture used:





The RNN model has yielded an accuracy of 95%.

**Results of the refined model:**

The model that we have built using CNN and RNN yields an accuracy of 95%. The model is able to predict the polarity of the amazon sentiment reviews accurately in most cases.

**Possible Refinements in future:**

Here are a few ways by which the accuracy of the model can be further improved:

Gated Recurrent Units(GRU):

GRU is a variation of recurrent neural networks (RNNs) that can effectively capture long-range dependencies in sequential data. Replacing LSTM with GRU in the model architecture might lead to better performance.

Transformers:

Transformers, particularly the Transformer architecture introduced in the "Attention is All You Need" paper, have shown remarkable success in various natural language processing tasks. Implementing a transformer-based model can potentially improve the model's understanding of context and relationships between words.

Bert Model:

Pre-trained BERT models have achieved state-of-the-art results in many NLP tasks. Fine-tuning a pre-trained BERT model on the sentiment analysis task can bring significant improvements to the accuracy.

**Conclusion and remarks:**

Our ongoing objective is to further enhance our model by leveraging advanced deep learning models. Nonetheless, it is essential to highlight our impressive achievement of 95% accuracy using CNN and RNN. We also recognize that there remains room for improvement and the need to explore additional facets while employing advanced techniques.