Steven Flores

Subash Chandra Pakhrin, Ph.D.

CS\_4317

February 24, 2023

**Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms**

The summary of Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms methodology that is used is collecting data of movie reviews from the IMDB. From there they clean and preprocess data by removing stop words and converting text to lowercase, stemming, and lemmatization. It also Converts the data into numerical features that are used by machine learning algorithms, such as bag-of-words, TF-IDF, and word embeddings. Then they choose a appropriate machine learning and deep learning algorithms, like as Naive Bayes, Support Vector Machines, Random Forest, and Recurrent Neural Networks. Then data is split into training models on training sets. The validation sets determine if the performance is validated. Hyperparameter tuning: Optimizing the hyperparameters of the selected models to improve their performance. Evaluating the performance of the models on the test set using metrics such as accuracy, precision, recall, and F1-score. Deploying the best-performing model for sentiment analysis on new, unseen movie reviews. It also involves collecting and preprocessing the gathered data and extracting relevant features such as selecting appropriate machine learning and deep learning algorithms, training, evaluating the models, and deploying the best-performing model for sentiment analysis on new movie reviews. With the machine learning being a GPT (Generative Pre-trained Transformer).

The novelty of the article depends on the specific approach that is made.: Novel methods of feature extraction that can develop to improve the accuracy of sentiment analysis on movie reviews. Combining multiple machine learning and deep learning algorithms in a hybrid approach could be a novel way to achieve higher accuracy in sentiment analysis. There is Developing domain-specific sentiment analysis models to movies or specific genres could be a novel approach to improve accuracy. Developing a real-time sentiment analysis system for movie reviews could be a novel contribution. This would involve using machine learning and deep learning algorithms that are efficient. As well as being scalable enough to handle large amounts of streaming data in real-time.

Word embedding is a technique used in natural language processing to represent words as vectors. Inside is real numbers in a high-dimensional space. The idea behind word embedding is to track and map words to a vector space in a way that captures the semantic relationships between them. This allows for more efficient and effective processing of natural language data in machine learning models. Then the word vectors can be used as input features for various tasks such as text classification, sentiment analysis, and language translation. Word embedding models are typically trained on large corpora of text using unsupervised learning algorithms. In the end this research can be a useful tool for understanding the sentiment of movie reviews, it is important to be aware of the limitations and potential biases of the models, and to interpret the results with caution.

In Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms is involved in using machine and deep learning techniques to predict the sentiment of movie reviews. There is also Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms typically focuses on analyzing the sentiment of individual person reviews, without looking at the broader context in which the reviews were written, like the reviewer's background, cultural context, and personal experiences. There can also be data faulty from sarcasm and irony in movie reviews, which can lead to incorrect predictions and reduced accuracy. This is due to machine learning and deep learning filtering the words and thinking the reviews are serious.

CNN 1D is a 1D Convolutional Neural Network is a type of deep learning network that is being used for processing sequential data like text data and time series data. It is a variation of the most common Convolutional Neural Network being the 2D CNN , which is used for image recognition. In a 2D CNN a 2D matrix of pixels is used for the image recognition and in a 1D CNN, the input is treated as a 1-dimensional matrix. The 1D convolutional layer performs convolutional operations and the input data along the time axis, allowing the network to capture local patterns in the data. As a result the output of the convolution operation is then passed through activation functions and pooling layers, and the process is looped many times to form a deep network. 1D CNNs are also used in applications such as voice recognition, natural language processing, and time series prediction. The results show it to be efficient in capturing local patterns in sequential data, which makes them well-suited for these types of problems.