**LITERATURE SURVEY**

**2.1 Introduction**

We go through various machine learning algorithms and choose the most suitable algorithm by referring various research papers and implement the most optimal algorithms and get the outputs through those processes.

**2.2 Literature survey**

2.2.1 CNN for different situations

In CNN for situations understanding based on sentiment analysis of twitter data by Shiyang Liaoa, Junbo Wangb, Ruiyun Yua, Koichi Satob, Zixue Cheng they proved CNN's ability to extract a set of features from a global dataset is the primary justification for using it in image analysis and classification. information, and it can take into account how these qualities relate to one another. The aforementioned method can increase the accuracy in classification and evaluation. Text data features can also be piecemeal collected for natural language processing in order to Take into account how these elements interact, but without taking into account context or the entire sentence, the sentiment can be misunderstood. We create a straightforward convolutional neural network model and evaluate it using the results that indicate that it performs more accurately in classifying Twitter sentiment than certain standard methods, including similar to the SVM and Naive Bayes techniques.

2.2.2 Sentiment analysis study for twitter

In sentiment analysis study for twitter using the various models of convolutional neural networks. Machine learning produces poor accuracy that is used in a larger range of applications. Consequently, the deep learning approach is being developed to increase sentiment analysis' accuracy. In order to increase accuracy performance, this study discusses several configuration options based on deep learning utilizing the Convolutional Neural Network (CNN) algorithm. To evaluate the effectiveness of the CNN models, a variety of parameters are offered, including the number of convolutional layers, the quantity of filters, and the size of the filters. The Word2Vec model for Indonesian has been utilized with the Indonesian-Sentiment-Analysis-Dataset, which consists of 10.806 tweets, as a word vector representation. The remaining 20% of the dataset is used for testing once the CNN models have been trained on 80% of it. Results from the proposed CNN models are compared and shown to be superior.

2.2.3 Twitter sentiment analysis using distributed word and sentence representation

In Twitter sentiment analysis using distributed word and sentence representation by Dwarampudi Mahidhar Reddy, Dr. N V Subba Reddy and Dr. Prema K V they introduced usage of Long Short-Term Memory (LSTM) Networks and Convolutional Neural Networks (CNNs) for Distributed Representation of words.Which lets us to capture Capturing local co-occurrence statistics and gives good performance with small (100-300) dimension vectors that are important for downstream tasks. It makes the process faster as only non-zero counts matter. Instead of utilizing conventional techniques or preparing text data, this research uses distributed representations of words and phrases. While the latter is used for the distributed representation of sentences, the first two are utilized for the distributed representation of words. This document has an accuracy rate of up to 81%. Out of the various techniques, it also offers the best and most effective approaches to generate distributed

2.2.4

In Using Word2Vec to Process Big Text Data by Long Ma and Yanqing Zhang they proposed a method to decrease the dimension of the feature vector which will be used to make our embedding layer for the cnn model which we prepare. Processing large data sets takes time since they can have a variety of distinct data types and sophisticated structures in addition to their large volume of data. If the learning algorithm can choose useful features or reduce the feature dimension, it will be more effective when taking the data dimension into account. Continuous Bag of Words (CBOW) and Skip-gram are the two learning models that make up Word2Vec. Word2Vec creates word vectors from text data that can be represented as a substantial passage of text or perhaps the complete article. In our work, we trained the data using a Word2Vec model and then assessed the degree of word similarity. In addition, we cluster the similar words together and use the generated clusters to fit into a new data dimension so that the data dimension is decreased which will decrease time consumption and increase performance.