

Sentiment Analysis Project:

1. We load Sentiment140 dataset with 1.6 million tweets and take 100K tweets for training and 30K for testing.
2. We Set target labels to be: 0 → Negative and 1 → Positive.
3. Then we preprocess the data to remove stop words, digits, special characters and punctuation, and return list of sentences each sentence consist of list of unique words.
4. We get word embedding of words from Pre-Trained FastText Model each word represents as feature vector with embedding dimension of 300.
5. Then we pass the embedding words to 3 CNN models to compare between them:
 - **First Model:**
the model architecture consists of multiple convolutional layers with varying kernel sizes (2, 3, and 5) and number of filters = 265, followed by batch normalization and global max-pooling layers. Then these layers concatenated together and add dense layer with sigmoid function.
This architecture allows the model to capture different n-gram features from the input text data efficiently. The input to the sequences of word embeddings with a fixed length of 17 and a dimensionality of 300.
This model gets 96.85% accuracy, 98% precision and 95% recall.
 - **Second Model:**
the model architecture consists of multiple convolutional layers with varying kernel sizes (2, 4, 6 and 10) and number of filters = 265, followed by batch normalization and global max-pooling layers. Then these layers concatenated together and add dense layer with sigmoid function.
This architecture allows the model to capture different n-gram features from the input text data efficiently. The input to the sequences of word embeddings with a fixed length of 17 and a dimensionality of 300.
This model gets 97.41% accuracy, 96% precision and 99% recall.

- **Third Model:**
VGG-16 Model
This model gets 88.81% accuracy, 92% precision and 85% recall.