

EARLY DETECTION OF DEPRESSION USING LINGUISTIC METADATA

Shruthi Ravishankar, Madhangi R, Apoorva NP

Introduction

- This system accounts for the **early** detection of depression using linguistic metadata.
- If depression is detected early enough, the victims can be urged to seek professional help.
- This project makes use of social media, especially Reddit to identify if the user is depressed or not.

Purpose

- The system can be used to predict if a user is depressed or not, to reduce the rate of suicides.

Dataset

- The Reddit Self-reported Depression Diagnosis (RSDD) dataset consists of Reddit posts for approximately 9,000 users who have claimed to have been diagnosed with depression ("diagnosed users") and approximately 107,000 matched control users .
 - All posts made to mental health-related subreddits or containing keywords related to depression were removed from the diagnosed users' data.
- The group of diagnosed users is made of users who
- have a post containing a high-precision diagnosis pattern (e.g., "I was diagnosed with") and a mention of depression,
 - do not match any exclusion conditions.

Evaluation Metrics

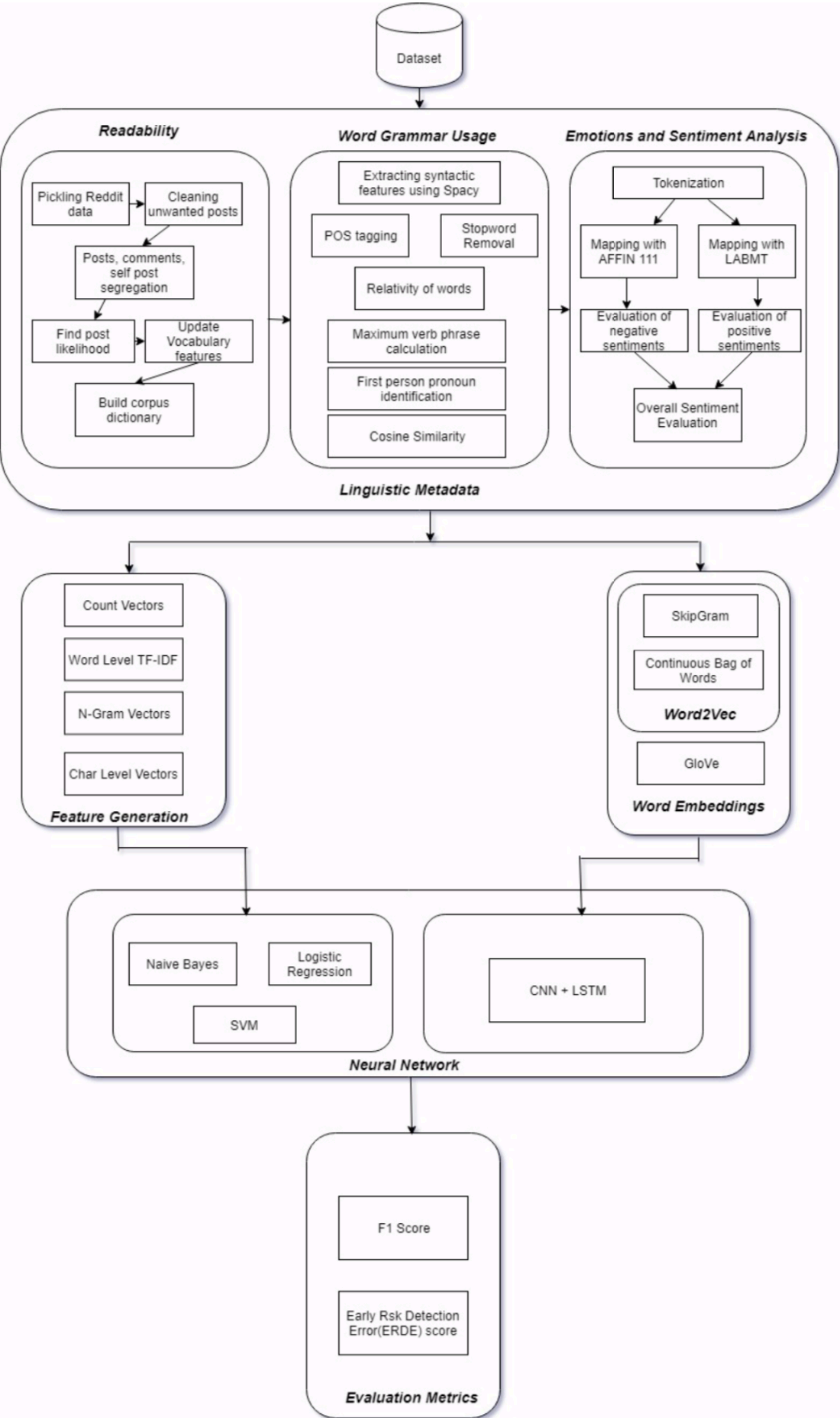
Early Risk Detection Error (ERDE) score :

- ERDE score is a measure used to reward early alerts.
- It considers the correctness of the decision and the delay taken by the system to make the decision.
- Used when the dataset is skewed towards the negative side.

Accuracy

- The accuracy is a measure of how accurate the obtained results are.

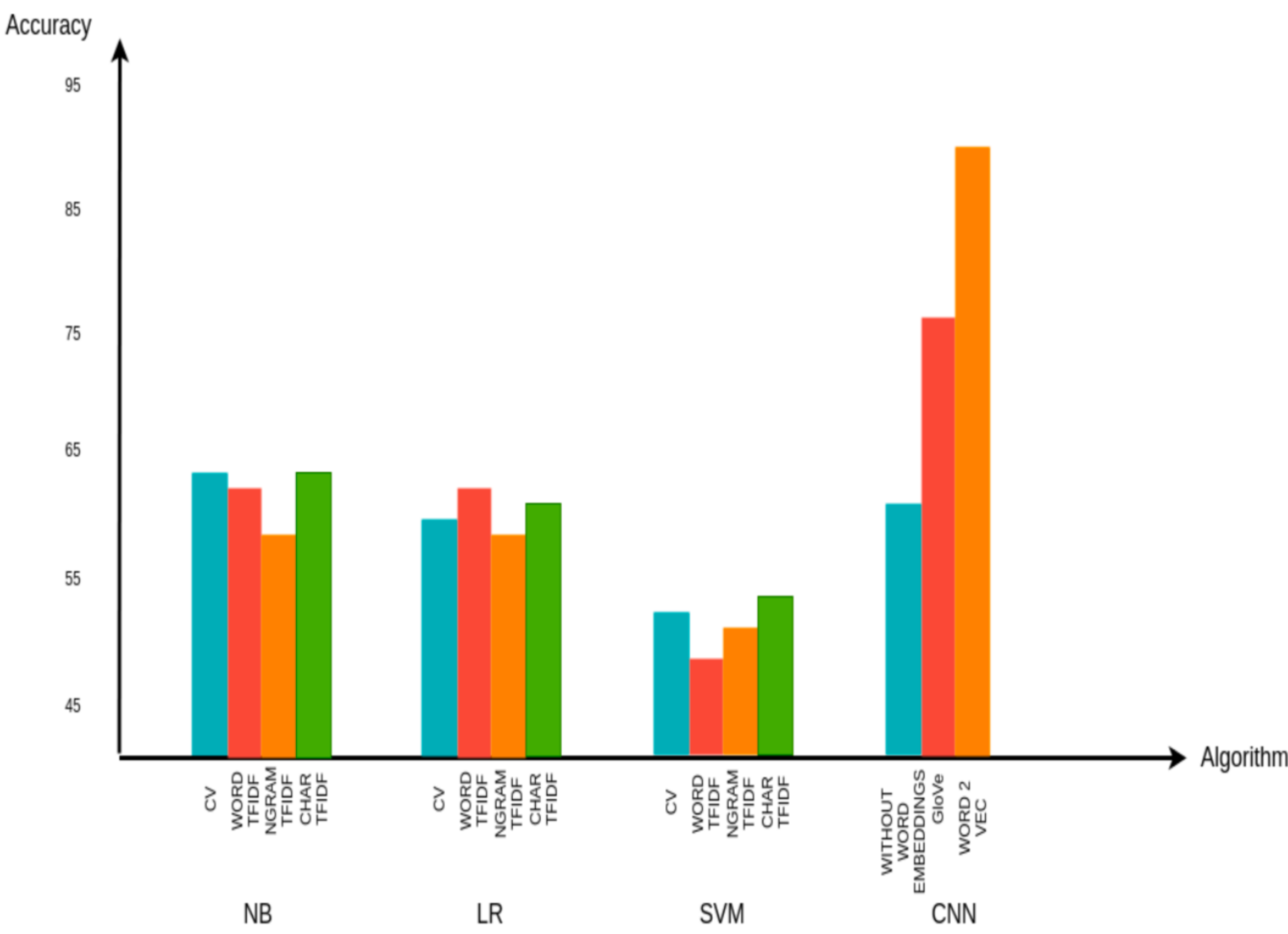
SYSTEM ARCHITECTURE



Results

Serial No.	Classifier	Feature	Accuracy	ERDE5
1	NB	CV	60.7	12.6
2	NB	Word-TF-IDF	61.3	12.9
3	NB	N-gram-TF-IDF	58.8	11.8
4	NB	Char-TF-IDF	63.1	13.5
5	LR	CV	59.5	12.07
6	LR	Word-TF-IDF	59.5	12.07
7	LR	N-gram-TF-IDF	62.9	13.6
8	LR	Char-TF-IDF	58.5	11.6
9	SVM	CV	52.5	10.5
10	SVM	Word-TF-IDF	49.3	10.4
11	SVM	N-gram-TF-IDF	49.6	10.7
12	SVM	Char-TF-IDF	51.5	11.2
13	LSTM+CNN	Without Word Embeddings	61.1	6.6
14	LSTM+CNN	GloVe	76.79	13.2
15	LSTM+CNN	Word2Vec	89.1	12.06

Comparison Chart



Discussion

- The future work can focus on using features such as voice and facial characteristics that can lead to better prediction rates.