**Thesis Presentation**

Thank you...

I am going to present the core part of our thesis the methodology part.

First of all, 5 methods we used are there. They are Linear SVM (Support Vector Machine), LSTM (Long Short-Term Memory), Bi-LSTM(Bidirectional), DT (Decision Tree), and LR (Logistic Regression).

To create the model and get the excellent raw data; system is divided into two phases. The first phase is the training phase where 80 % of work has been done and another phase is the testing phase where 20% of work has been done.

The whole process is divided into four sequential phases to get a better understanding. They are Dataset preparation --> where all the data has been collected and filtered. The next phase is the Pre-processing --> Here there are 7 works like lower case converter, bad symbol, emoticon, stop-word remover, tokenization, and streaming.

**Next slide, please...**

The next phase is Feature Selection and Extraction.

In this part, some work like parts of speech tagging, gram features like Uni/Bini/ Tri has been done. There are TF-IDF = term frequency-inverse document frequency and word2vec also been done.

The last and final phase is for the methods we have used in our thesis. They are Linear SVM, LSTM, Bi-LSTM, Decision tree, and Logistics Regression.

**Next slide, please...**

This slide is the flowchart of the work.

As I mentioned earlier about the system is divided into two phases. We can see on the left side the Architecture of the training phase. Here the labeled data enter the training phase and then the data preprocessing part comes where lower-case converter, emoticon, stop-word remover all those works happen. Then after passing features selection of pos tagging, uni, bi, tri, and word2vec it goes to the classifier and finally predicted hate speech label enter into the evaluation and got the performance.

On the right side, it is the testing phase work. Here the leveled data enter the testing phase and pass to the predictive model. Here all 5 methods like the SVM, LSTM, etc works. And it passes to the predicted speech level. Some predefine hate text also goes to the evaluation with this. And finally, the evaluation gives the outcome of performance accuracy.

**Next slide, please.**

This slide is the system overview which is the combination of the Training phase and Testing phase.

Here the collection of data goes to the data processing part where lower case converter, emoticon, stop-word, punctuation remover, tokenization, and streaming works. After that, the clean data go to feature selection to split in some words. After data splitting, some data go to the test phase, and some go to the training phase. Train data pass the model construction and train the model directly go to the model evaluation. In evaluation, the prediction using data and evaluation of model show the result performance in accuracy term.

**Next slide, please......**

This slide is built with dataset information. we use 26554 sentences and 265543 words for implementation. Mainly we collected data online and create 200 sentences by ourselves. We can see it is separated into two phases named training and testing phase with detecting four hate emotions named stupidity, pathetic, greedy, and retarded. Seeing all the information the stupidity class gives the highest amount from the data set which is 10033 from training and 3202 from the testing phase. And the retarded class gives the lowest amount that is 1215 from training and 805 from the testing phase.

**Next slide, please….**

This slide is the sample data.

From our dataset, we have picked a random line here. And shown the gram features. So, the sample text is it ain't nothing to cut a bitch off. So, if we do the unigram feature it will take a single word at a time. So, it will be like it, ain't, nothing and so on. If we do the bi gram feature then it takes two words at a time like it ain't, ain't nothing, nothing to, and so on. And finally, if we chose the Trigram feature then it will take 3 words at a time and show the result of it ain't nothing, nothing to cut, cut a bitch, and so on. That's all from my side. My next presenter will continue now.