

# Meeting - 6

**4 March 2020 / 7:00PM/BIG DATA ANALYTICS LAB,CC3**

## **Attendees**

Amit Kumar ,Project Instructor  
Harsh Goyal (IIT2018114)  
Aaditya Gadhave (IIT2018144)  
Sourabh Gupta (IIT2018149)  
Meet Singh Gambhir (IIT2018158)  
Tushar Atrey (IIT2018159)

## **Agenda**

1. Main agenda of this meeting was topic modeling.
2. Getting familiar with Gensim , LDA++ and Maillet tools used for topic modeling
3. Introduction to WEKA and ARFAA file
4. Making the corpus and training the data

## **Notes**

- The meeting started with the briefing of the work we did earlier
- The instructor told us how to make a corpus for the topic modeling tools.
- Instructor told us to divide the data into two parts :
  - I) training data (70%)
  - II) testing data (30%)
- Instructor told us how to train our data and how to test whether it is working correct or not
- Instructor assigned the task i.e to make the table in which rows will be the pairs of bugs and in column there will be different kinds of paths(mentioned in the meeting) . Each index value means the competitive measures between a bug pair.
- We have to make such four matrix because there are four methods of competitive measures(path count,normalized path count,random walk, symmetric random walk)

- Tasks to be completed until the next meeting were assigned

**Action Items**

1. Team is required to do the topic modeling
2. Team is required to train the data.
3. Team is required to make tables for bug pair and competitive measures.