

# Project Outline

## Introduction to NLP

**Team CheeseMaggi:** Palash Sharma (2019101082), Aryan Jain (2019101056),  
Arth Raj (2019101094)

---

### 1. Problem Statement

Preparing named entity tagged data by utilising existing annotated knowledge using snorkel so as to achieve Named Entity Recognition (NER).

### 2. Implementation Flow

1. **Dataset understanding and topic research:** Choosing and understanding the relevant details of the dataset chosen and performing thorough reading of techniques used for NER.
2. **Preliminary NER system:** Etch out labelling rules and implement them using Snorkel to obtain Named Entity tagged data.
3. **Analysis:** Assessing the quality of the data procured from the above step based on the several different parameters chosen appropriately.
4. **Model improvement:** Bootstrap a Transfer Learning Model for improving the F1 score of the data.

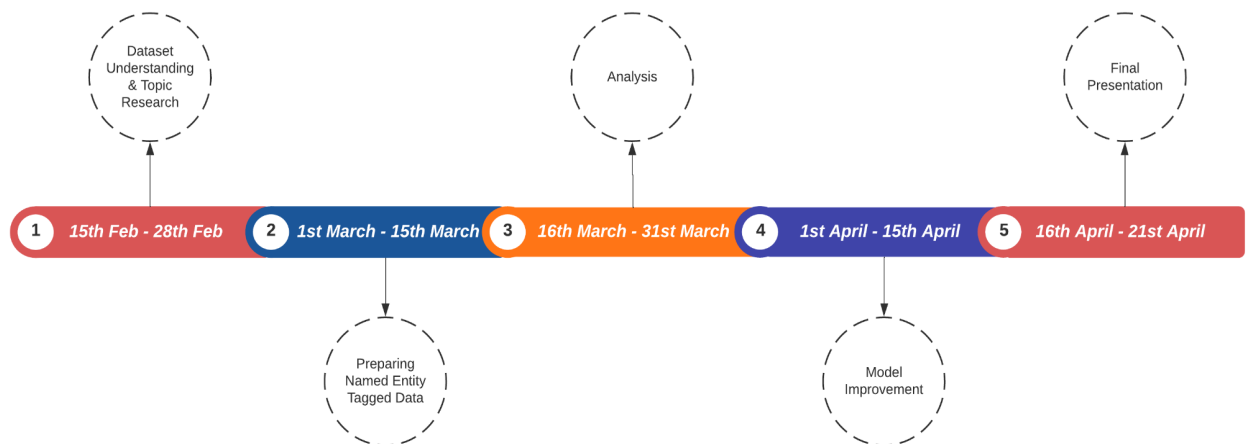
### 3. Deliverables

1. **Interim Deliverable:** Initial implementation of NER system using Snorkel along with a detailed report of the implementation, challenges faced and the plan ahead.
2. **Final Deliverable:** Complete system for NER with a boosted F1 Score along with a report that mentions the details of implementation, novelties incorporated (if any), complete analysis of the model with various industry standard NER systems and benchmarks and our learnings from the project.

## 4. Dataset

1. <https://huggingface.co/datasets/conll2003> - Conll2003 is a widely used dataset with more than 20k sentences broken into tokens.
2. <https://github.com/juand-r/entity-recognition-datasets> - Has a wide variety of languages and is used for various NLP tasks including NER.

## 5. Timeline



## 6. References

### 6.1 Online Links referred

1. <https://www.snorkel.org/>
2. <https://www.snorkel.org/use-cases/01-spam-tutorial>
3. <https://www.analyticsvidhya.com/blog/2021/06/part-10-step-by-step-guide-to-master-nlp-named-entity-recognition/>

### 6.2 Research Papers referred

1. <https://arxiv.org/pdf/2004.14723.pdf>
2. <https://ojs.aaai.org/index.php/AAAI/article/view/6009>
3. <https://www.jbe-platform.com/content/journals/10.1075/li.30.1.03nad>