* **Introduction (3-5 sentences)**

**We are attempting to build a system to detect Fake News. Now a days this is the major concern due to the rapid spread of news through different mediums of communication and misleading many of us. If there is a system which can detect the fake news it will help us to deliver the rightful news and minimize the work load to editors who are doing this job manually.**

* **­Related Work (10-30 sentences)**

We read many different research papers and studied a variety of approaches to solve the problem. The Fake News challenge has been attempted by many people and there was a wide range of research papers, articles and blogs on the topic. We have gone through Fake News Challenge repository which has a baseline implementation provided by the organizers. It works almost perfectly on identifying the related/unrelated relationship for the pairs. There are different approaches like using recurrent neural network models with memory enabled units like LSTMs and conditional encoding LSTMs with attention. Besides that using bag-of-words followed by a three-layer multilayer perceptron ( BoW MLP ) provided better results amongst all.

* **Project Topic and Proposed Solution (10-30 sentences)**

**The title of the project is “Fake News Detection”.**

**The Dataset we are going to use is from** <https://github.com/FakeNewsChallenge/fnc-1> **which contain two files**

1. **Id, Article body**
2. **Headline, Article Body Id, Stance (unrelated, related, agrees, discuss)**

**We are going to classify the news into four stances based on the relevance between headline and the body.**

**Different steps involved for this procedure are as follows:**

* **Preprocessing datasets using NLP techniques.**
* **Finding correlation between headline and body.**
* Extracting Refuting words from dataset
* Word N-grams extracted from the headline and body pair and finding the overlapping grams
* Building Machine learning model to classify stance to one of the above four mentioned
* Performance is measured by giving 75% weightage for classifying correctly on of the three in related stance and 25% weightage to classifying unrelated/related.
* The final result will be the classification of test data into 4 stances mentioned above.
* **Project Timeline (fill in this table, no additional text required)**

|  |  |  |
| --- | --- | --- |
|  | Date | Milestone achieved |
| 1 | Sept 7th | Project Proposal submitted |
| 2 | Sept 18th | Reference implementation |
| 3 | Sept 30th | Project at least 50% done, project progress report due |
| 4 | Oct 15th | Comparing different ML models for the problem |
| 5 | Oct 30th | Implementation of the chosen model |
| 6 | Nov 15th | Accuracy Metrics and Preparation of Final Report |
| .. | Nov 25th | Project Complete, final report due |

* **Team Roles and Contributions (fill in the table, no additional text required)**

|  |  |  |
| --- | --- | --- |
|  | Team Member Name (800 id) | Responsible For |
| 1 | Sai Supreeth Segu(801075915) | Literature Study, Pre-Processing, implementing ML Algorithm1, Writing Report |
| 2 | Narendra Kumar Vankayala(801081957) | Literature Study, Correlation, implementing ML Algorithm2, Accuracy Metrics, Writing Report, |

* **References (provide any background references, mandatory section)**

1. http://www.fakenewschallenge.org/
2. Mrowca, D., Wang, E., & Kosson, A. Stance Detection for Fake News Identification.
3. Xiaowei Wu, Sizhu Cheng & Zixian Chai for Fake News Stance Detection
4. Chaudhry, A. K., Baker, D., & Thun-Hohenstein, P. Stance Detection for the Fake News Challenge: Identifying Textual Relationships with Deep Neural Nets.