# Capstone Project Proposal Report
# (Individual Report)

**Instructions:**

This form is to be completed by each student doing Project registration to fulfill their senior design or capstone requirement. It must be completed and submitted to your Guide. Each student must complete this form individually.

This report is to be completed during the starting of the semester, while the project description report will be completed during end of the semester.

| | | |
|---|---|---|
| Guide Approval (initials/date): | DR. D. SUMATHI | 14/06/2020 |

## CAP4001– Capstone Project Proposal Report

| | |
|---|---|
| **Student Name** | YALAMADDI ABHINAV |
| **Student Register Number** | 17MIS7077 |
| **Program** | Integrated M. Tech Software Engineering |
| **Semester/Year** | Sem-10 / 5th year |
| **Guide(s)** | Dr. D. Sumathi |
| **Project Title** | Abstractive Text Summarization |

**Team Composition:** Provide the information below for each member of the **project team**. Include **all** project team members, not just those in your discipline or those enrolled for Capstone project. Please also include yourself!

| **Reg. No** | **Name** | **Major** | **Specialization** |
|---|---|---|---|
| 17MIS7077 | YALAMADDI ABHINAV | Computer Science | Software Engineering |

**Project and Task Description**: Provide a brief (one or two page) technical description of the design project and your specific tasks, as outlined below: (use a separate sheet)

(a) Provide a summary of the project, including a description of the project and its requirements, the purpose, specifications, and a summary of the approach. If this is a continuing project, you may use and/or edit the same project description.

Abstractive text summarization is the task of generating a headline or a summary of a few sentences that captures the main points of an article or passage. We use the adjective 'abstractive' to describe a summary that is more than just a few existing passages or sentences extracted from the source, but a compressed paraphrasing of the main contents of the document, potentially using vocabulary not found in the source document.

The purpose of this project is to mimic the basic principles in this study of state-of-the-art abstractive text repeating models to investigate various processes for attention until I had a decent working base. I had inspired by many research articles that have obtained outstanding achievements by incorporating several new features. This research is being carried out with three rounds, improving the word embedding, the complexity of the encoder decoder and attention. The third model also introduces a bilinear attention mechanism that increases the rate of loss of training.

Hardware requirements:

● OS: Windows 7 or above, Unix

● RAM: 8GB (recommended)

● Storage: 10GB (minimum)

Software Requirements:

● Python (3.7 above)

(b) Describe the specific role and tasks that **you individually** will be completing as part of the design of the project. What **specific deliverables** will you produce?

- To design an efficient system which can produce a summarized form of a sentence in human readable form.
- A Three Layer Stacked LSTM Encoder-Decoder model with Global Attention Mechanism was used during implementation. On the training set, I was able to achieve an accuracy of 77.27 percent using this model (constituting 80 percent of the dataset). This model also achieved a cumulative BLEU-4 score of 0.8800 on the test set.
- At the completion of the project, it was able to produce a summarized text output with better accuracy.

(c) Discuss in detail the specific approach that will be used to complete **your** portion of the design.

As stated, before this project is implemented using 3 models as a stacked layer.

1. The first model was a basic one-way LSTM encoder decoder with randomly initialized words. As mentioned later, discovered that using preexisting word embedding would be better, and hence iterated Model 2 using this model.

2. The model with a two-way LSTM encoder and a unidirectional LSTM decoder has been implemented.

3. Implementation of a two-way LSTM encoder in the last model and incorporated the LSTM decoder worldwide attention. Previously provided attention scoring functions. The final secret state (forward and reverse) of the encoder is linked and utilized as the initial hidden state of the decoder.

(d) Describe the phases of the design process that will be incorporated and what work will be accomplished during those phases. (You may attach a Gantt Chart)

| Stage | Week | Task |
|---|---|---|
| Code | 1 | Define Text Summarization |
| | | Resource allocation and related works |
| | 2 | Algorithm implementation |
| | | Data cleaning |
| | 3 | Algorithm enhancement |
| | 4 | Code review |
| | | Benchmarking with other algorithms |
| QA | 5 | Record and documentation |
| | | Generating report analytics |
| | 6 | General Algorithm enhancements |
| | | Grooming bugs |
| Release | 7 | Fixing bugs and code review |
| | 8 | Review and grooming the code |
| | | Documentation |
| | | Final Code Review(possibility of enhancement) |

**Outcome Matrix:** Describe your plan to demonstrate each of the outcomes below.

| **Outcomes:** | **Plan for demonstrating outcome:** |
|---|---|
| a) an ability to apply knowledge of mathematics, science, and engineering | I deploy a model based on my experience over the past four years, and this project covers all areas of expertise, and I will do my best to implement it. |
| b) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability | This project will be using only software resources for the most essential and hardware resources, such as a laptop with a decent GPU power processor or, in general, a good graphic card. It is consistent with all restrictions and is carefully maintained. |

| | |
|---|---|
| c) an ability to function on multidisciplinary teams | With the support of my project guide Dr. D. Sumathi, I will carry out my project independently and deliver an overview of what I described in my objectives. |
| d) an ability to identify, formulate, and solve engineering problems | Encountering numerous challenges in all phases and researching the topic with my own interest and finding a way of dealing with it as soon as possible. |
| e) an ability to communicate effectively | I can convey my proposal very clearly since I have a clear understanding of what I am trying to do. |
| f) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice | To accomplish this project, I'm able to resolve the technical challenges through the implementation of current engineering methods by implementing all the essential standard tools. |

**Realistic Constraints:**

This has been the first experience in establishing a large-scale data project, and the amount of learning that I have done is incomparable from any other project. The creation of every starting code was a significant problem. The need to grasp what each functions call accomplished, but to rely on the code of earlier.

Few Constraints are as:

- Building scripts to analyze large volumes of text.
- Summarized text should be trimmed in a such a way that the sentence formed should be meaningful.

**Engineering Standards:**

Discuss the Engineering Standards that will be followed and maintained in the Project:

- Usage of recognized software like: Anaconda Navigator and Google Collab
- For coding: Jupyter Notebook which is a well-known editor for Python language
- Remote connection through Command prompt has been established to maintain a secure connection through Anaconda Navigator.