**IS597MLC: Final Project Proposal**

**NetID:**

**Student Name:**

**Title**

Provide a brief and descriptive title for your project.

**Motivation & Objective**

Describe the background information of the topic you selected for your final project. Also, provide detailed explanation of your motivation and objective. Include 2-3 research questions you intend to answer throughout this project. This section should be described in 2-3 paragraphs.

**Related Articles**

Provide a citation of 4 scientific articles you selected to include in your literature review. And briefly describe each article in 3-4 sentences.

**Data**

1. **Data Collection**

Explain in one paragraph (6-8 sentences) about how you collected your data and what your data set looks like. For example, a source of your dataset, size of data set, the total number of records available (how many instances and attributes it contains), data format in the dataset (e.g., column names and brief explanation), etc. Make sure that the number of instances (rows) in your data set is not less than 30,000 given the model training requires a big size of data.

If your data set contains multiple files, you will need to merge them into one combined file or load data separately then process them. Describe your approach in detail. You will also need to identify the target class that you are planning to use as a label for training and testing your models (note that we are working on supervised machine learning). Provide a table that describes the attributes in your dataset. Also, you might want to include an image or screenshot of a page of your data set.

|  |  |
| --- | --- |
| Attribute Name | Description |
|  |  |
|  |  |
|  |  |

1. **Data Pre-processing**

Explain in 1-2 paragraphs about how you will plan to clean your data, such as what steps to be used, what packages & libraries to use, etc. You should answer questions, e.g., How will you handle duplicates, missing values, etc.? What attribute or column have you determined to use for your label class (can directly use it as it is or need to convert into a new target class using your own category scheme)?

**Analysis & Methodology**

Explain how you plan to achieve your goal in 2-3 paragraphs. What techniques of analysis do you want to use to analyze your data? e.g., TF-IDF, POS Tagging, One-hot coding, etc. What algorithms do you want to apply for training models? Do you have any evaluation metrics in mind to assess the performance of your trained models? Note that you do not have to stick with the methods or steps you describe in your proposal as you make further progress. You can change them and explain the rationale in your final report.

**References**

Please insert citations of any scientific articles you include in this proposal to support your research plan. Use standard citation styles such as APA, MLA, Chicago, etc.