AML 2304 2, FINAL PROJECT PROPOSAL

PROJECT TITLE	Predicting Success in Climate Change Posts, Part II
SUBMITTED BY	Mary Gomez
DATE OF SUBMISSION	11/22/2023

First Name	Last Name	Student Number
Mary	Gomez	C0891136
Rehan	Khokhar	C0896278
Andres	Correal	C0872634
Viki	Patel	C0906295
Bhavya	Vadher	C0894977
Prashant	Bhattarai	C0898956

MOTIVATION	Climate change is an issue that needs attention and action. The best way to reach as many people as possible is by using social media. Therefore, understanding and predicting the success of climate change posts will help current and future leaders who want to influence public action on this matter. In **part I** of this project, we aimed to uncover what elements make a post on climate change thrive and generate attention from the public. Additionally, we had a first overview of the potential of predicting post-success by training two machine learning modes. As we saw an excellent potential to achieve a good performance in model prediction, in this **part II,** we will use several strategies to optimize the model performance (Accuracy).
METHOD	• Additional of the data will be taken from Reddit API (https://www.reddit.com/dev/api). • We will use primarily nltk, scikit-learn and Hugging Face to conduct our experiments.
INTENDED EXPERIMENTS	• Use a feature selection technique to determine the best predictors • Feature Engineer new columns • From the textual data, use the top words, bigrams, trigrams as features of the model. • Use stemming or lemmatization to reduce dimensionality • Explore the use of NLP (Natural Language Processing) vector embeddings from Hugging Face for textual data into features. • Test and compare several machine learning models • Optimize the model performance by doing hyperparameter tunning using GridSearch technique.
PLANNING AND MILESTONES	• Get the best predictors - Prashant • Feature Engineer new columns using learnings from part I – Rehan • Choose bigrams and trigrams as features of the model – Viki • Compare machine learning models- Andres • Compare vectorization methods – Bhavya • Hyperparameter tunning - Mary • Reporting - Mary