**CSCE 5290: Natural Language Processing Project Proposal**

**Project Title:**

Autonomous Tagging of Stack Overflow Questions

**Team:**

1. Aditya Pujari (11491374)

2. Sai Tarun Gunda (11516657)

3. Nithish Reddy Boddhi Reddy (11555383)

4. Gagan Sai Ram Anvesh Achanta (11447940)

**Goals and Objectives:**

**a) Motivation:**

In today’s Programming world, one of the best ways to discuss about the programming related errors/ issue would be Stack Overflow. If a user posts his question in Stack Overflow, he must manually tag the related topic he is looking for to get the issue resolved/ redirected to the correct community of people who are looking/ already working in that field. For example, if we are looking for some machine learning topic and have a query related to some training of SVM model after posting the question we have tag it as Machine learning, so that it reaches the appropriate set of people. Instead of tagging Machine learning if we tag Accounting it would be helpful to resolve the problem.

**b) Significance:**

If we can develop a model where it automatically tags the question with its respective task i.e., it automatically tags a machine learning problem to machine learning tag and accounting questions to accounting tag and so on. By doing so the questions we will get answers/solutions quickly as it reaches the right set of people.

**c) Objectives:**

Our main goal is to develop a model where it can auto tag the questions posted in the Stack overflow. We developed a multi classification system that gives tags to discussion forums automatically. We use a dataset of Stack Overflow questions to construct and test our classifier.

**d) Features:**

We are planning to find n-grams and count the total frequency of words eliminate the most common words like ‘and’,’or’,’between’, so on which do not add any meaning to the sentence classification. We are planning to use Term Frequency Inverse Document frequencies to reweight the count on how it is adds meaning/ relevant to the topic we want to classify.

References:

1.<https://www.kaggle.com/competitions/multilabel-bird-species-classification-nips2013/data>

2. <https://towardsdatascience.com/journey-to-the-center-of-multi-label-classification-384c40229bff>

3. Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. CoRR, abs/1301.3781

4. Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. CoRR, abs/1301.3781.

**Git Hub :**

https://github.com/Gagan-Achanta/CSCE\_5290\_NLP\_Project.git