PROJECT PROPOSAL

TAG PREDICTION OF STACK OVERFLOW QUESTIONS USING ML ALGORITHMS

Introduction

Currently, with huge amounts of data present throughout the world, technology is growing at an enormous pace and is accessible to everyone. There is a huge demand for filtering and sorting the data based on a person's need. One such platform, StackOverflow, is a place which is used in finding a solution to Engineer's daily faced problems which are taken up by a large community of Engineers, students and tech enthusiasts. To ease up this process at both ends, filtering the questions based on requester's needs is really important, so as to make the process more efficient. We wish to work on this and provide an automatic tag predictor for StackOverflow posts.

Datasets

- ★ Facebook Recruiting III - Keyword Extraction | Kaggle
- ★ Predicting tags for StackOverflow | Kaggle

Tentative Progression

1. **Data Preparation**
2. **EDA : Exploratory Data Analysis**
3. **Implementing Algorithms**

   Being a team of four members, we plan to rigorously implement 4 classification algorithms and comparatively analyze them. The algorithms would be from the subset of Naive Bayes, SVM, Decision Trees, Multi Layer Perceptron, SDG Classifier, etc, specific algorithms would be subject to further discussion with the instructor and TA's.
4. **Tuning and Analysis**

   The algorithms will be tuned with the help of different hyperparameters and cross validation techniques. This will help us gain insight into the bias-variance tradeoff, the theoretical foundations of which were laid in class lectures.
5. **Results and Conclusion**
6. **Implementation of the state of the art model**

   As the problem is already well established and work has been done on it, If time permits, we would like to study and implement the current existing SOTA model for this problem by reading through the latest implementations and literature work on it

Please Note : The problem of tag classification has been studied fairly rigorously. We plan to systematically apply and analyze some of the methods learned in class, to gain an analytical insight of the hands-on implementation of these various techniques.

TEAM DETAILS

Name	Roll No.	Regular/Audit
Aditya Pande	22M2108	Regular
Kasuba Badri Vishal	22M2119	Regular
Prathamesh S. Yeole	22M0795	Regular
Sourav Paul	22M2113	Audit