1)Problem statement –

CodeForce is a Programming language Competitive platform where Questions are posted , usually its quite difficult to come up with a approach for solving looking at the description . So the tagging for questions is important for a better user-experience. Tags may be generic like ‘math’ or more specific like ‘Brute Force’.

In this machine learning problem , the final goal is to predict the tag based on the textual description of the Question which is abbreviated as ‘problem statement’ and also optimize their precision and recall values for [[CodeForce](https://codeforces.com/)] platform. Each problem has multiple tags associated with it and is a Multi-label classification.

We are planning to train the data using catboost , Multi-label version of KNN ,LSTM with one-hot encoding , tf-idf decision tree and Logistic Regression. But we are still researching on which model to use for training . Based on the frequency of tag occurrences we are planning to go for either micro-average f1 score or macro-average f1 score . We are also researching to show the misclassification of tags using hamming loss calculation.

At the moment we are researching to store the csv data in the sqlite db through python framework as well , if the dataset is large .

Reference: https://arxiv.org/pdf/1911.12224.pdf

2) Description of data set –

We are referring to the research paper to implement the ML , but we got the dataset from the kaggle.

<https://www.kaggle.com/immortal3/codeforces-dataset?select=data.csv>

3) Implementation plan - this section briefly describes the tentative plan for implementation, milestones and timeline .

|  |  |  |
| --- | --- | --- |
| SL No | Project Phases | Week |
| 1 | Data Preparation and Pre-processing | Week1 |
| 2 | Explorary Data -Analysis | Week2 |
| 3 | Model Selection and Training | Week3 |
| 4 | Performance Evaluation of Model and Result | Week4 |
| 5 | Report Generation | Week5 |

4) Team members & task allocation - this section list names of all team members and defines tasks for each member.

|  |  |  |  |
| --- | --- | --- | --- |
| Sl No | Project Phase | Team Members | Week1 |
| 1 | 1 | Shivani Mogili  Deepshika Reddy |
| 2 | 2 | Amit  Shivani | Week2 |
| 3 | 3 | Amit  Shivani  Deepshika | Week3 |
| 4 | 4 | Amit  Deepshika  Shivani | Week4 |
| 5 | 5 | Amit  Deepshika  Shivani | Week4 |