

Multi-label Classification Tag Prediction in the Context of Programming Challenges

Problem statement

CodeForce is a Programming language Competitive platform where Questions are posted, usually its quite challenging to come up with an approach for solving looking at the description. So, tagging for questions is essential for a better user experience. In this machine learning problem, the final goal is to predict the tag based on the Question's textual description, abbreviated as 'problem statement', and optimize their precision and recall values for the [CodeForce] platform. Each problem has multiple tags associated with it and is a multi-label classification.

Description of data set

We refer 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>

Data will consist of problem statement and problem tags columns.

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 | Week 1 (27 Oct) |
| 2 | Exploratory Data Analysis | Week 2 (3 Nov) |
| 3 | Model Selection and Training | Week 4 (10 Nov) |
| 4 | Model Implementation, Performance Evaluation of Model and Result | Week 4,5 (17 Nov) |
| 5 | Final Report and Presentation | Week 6 (1 Dec) |

Team members & task allocation

| Project Phase | Team Members |
|--|---|
| Data Preparation and Pre-processing | Shivani Mogili Deepshika Reddy |
| Exploratory Data Analysis | Shivani Mogili Amit Singh |
| Model Selection and Training | Shivani Mogili Amit Singh Deepshika Reddy |
| Model Implementation, Performance Evaluation of Model and Result | Amit Singh Deepshika Reddy Shivani Mogili |
| Final Report and Presentation | Amit Singh Deepshika Reddy Shivani Mogili |