

StackOverflow Tag Predictor

Machine Learning Project Proposal

Navya Aggarwal
2018349

Nitin Gupta
2018251

Sandeep Kumar Singh
2018363

1. Motivation

Stack Overflow is the largest, most trusted online community for developers, students, and other educationists. It allows users to ask questions, answer questions, vote questions, and answers up or down and post questions and answers in a similar manner to the Wiki or Quora. StackOverflow allows users to manually assign between one to five tags to a posting. The tagging feature is well used by experienced and platform-known users, but new and inexperienced users tend to miss appropriate tags. A new or inexperienced user can miss an important appropriate tag or use an inappropriate task. In this project, we are developing a predictor that is able to assign tags based on the content of a question.

2. Related Work

The tags for a question are a very important aspect of the questions, as it is used by the website to optimize search. So, the use of the correct and appropriate tag is a must. To tackle this problem of tag prediction, previously, papers and models are built by people around the globe. During the course of our project, we plan to go through some of the papers and articles. The link of all the relevant papers and articles that we came through are listed below in the references section.

3. Timeline

Week 1-2:	Data Collection
Week 3:	Pre-processing and Data Visualization
Week 4:	Feature Extraction
Week 5:	Analysis of features
Week 6-10:	Explore different machine learning algorithm
Week 11:	Checking for inaccuracy in predictions(over-fitting, under-fitting)
Week 12:	Report writing

4. Individual Tasks

Team	Team Member/s
Data Collection	Navya and Nitin
Pre-processing and Data Visualization	Sandeep and Nitin
Feature Extraction	Navya and Sandeep
Analysis of Features	Navya and Nitin
Explore Different Machine Learning algorithms	Navya, Nitin, Sandeep
Checking for inaccuracy in predictions (over-fitting, under-fitting)	Sandeep, Nitin, Navya
Report Writing	Sandeep

5. Final Outcome

The final objective is to train a model such that, it can provide up to five tags for a given question from StackOverflow. We will use different algorithms to pick-up insights from the question title and question body, and using the model we will predict the tags for the question.

In the future, we can use this model to predict tags for different questions, and we can also make a website, to make tags prediction to everyone.

References

- [1] **Predicting tags for stack overflow questions using different classifiers** - https://nrega.nic.in/netnrega/circular_new.aspx
- [2] **Predicting Tags for StackOverflow Questions** - <http://cs229.stanford.edu/proj2013/SchusterZhuCheng-PredictingTagsforStackOverflowQuestions.pdf>
- [3] **Predicting tags for the question in StackOverflow** - <https://medium.com/datadriveninvestor/predicting-tags-for-the-questions-in-stack-overflow-29438367261e>