

# Autonomous Tagging of Stack Overflow Questions

CSCE 5290: Natural Language Processing, Spring 2022

**Instructor:** *Dr Sayed Khushal Shah* ([sayed.shah@unt.edu](mailto:sayed.shah@unt.edu))

**Team:** Alekhya Vachakarla ([alekhyavachakarla@my.unt.edu](mailto:alekhyavachakarla@my.unt.edu))

**Github Link:**

<https://github.com/AlekhyaVachakarla1225/Autonomous-tagging-stackoverflow-questions.git>

## Introduction

Online responsive discussions, for example, Stack Exchange and Quora are turning into an undeniably well known asset for training. Key to the usefulness of a large number of these discussions is the idea of labeling, by which a client names his/her post with a proper arrangement of subjects that portray the post, to such an extent that it is more effortlessly recovered and coordinated. The quantity of data is increasing day by day on these forums but there is no productive or automated way to classify the data currently. We propose a classification that naturally labels clients inquiries to upgrade client experience.

## Motivation

Stack Overflow is the biggest platform where different people like developers, students, designers etc. use to share, learn and develop their skills and knowledge and build their careers. Most of them use stack overflow while developing a technical feature or to resolve any issues or errors in their code/programs. Every month, more than 50 million engineers come to Stack Overflow to learn and share their insight. The platform allows the users to ask and answer the questions at same time. Many people also vote and edit answers.

## Significance

- Why is Autonomous tagging of stack overflow questions Important?

A question posted on stack overflow contains three different segments title, description and tags/labels. By using the title and description, the system should be able to automatically suggest the label related to the subject posted on the website. These labels are very important for the efficient working of the stack overflow platform.

- Tagging of the questions is specifically useful for indexing data based on the tags.
- If the questions posted are not segregated or grouped correctly, then the questions will lose all sense of direction in a pool of un-addressed questions.
- Assigning labels to the questions will also help in clubbing the similar questions.
- Assigning tags/labels correctly to the questions will make it more likely to reach the proper audience and thus there will be a higher opportunity in getting the question answered.

- Incorrect labeling of the questions will also impact the experience of the users on stack overflow.

## Objective & Features

The objective of my project is to build different models which classify the category of the question posted or inputted. With the help of deep learning models and bert models we can implement this mechanism with more precision and accuracy

- Using different NLP techniques to prepare and process the data.
- Creating a Fully trained model and training the model with the dataset
- Evaluate the performance of the model and check whether the model is able to determine the category of the given input text correctly or not.

## References

- Mihail Eric, Ana Klimovic, Victor Zhong. Autonomous Tagging of Stack Overflow Questions. Paper published on December 8, 2014 in stanford edu.  
<http://cs229.stanford.edu/proj2014/Mihail%20Eric,%20Ana%20Klimovic,%20Victor%20Zhong,MLNLP-Autonomous%20Tagging%20Of%20Stack%20Overflow%20Posts.pdf>
- Virik Jain, Jash Lodhavia. Automatic Question Tagging using k-Nearest

Neighbors and Random Forest. 2020 International Conference on  
Intelligent Systems and Computer Vision

[10.1109/ISCV49265.2020.9204309](https://doi.org/10.1109/ISCV49265.2020.9204309)

- Meghashyam Chinta, Predicting Tags for the Questions in Stack Overflow

[https://medium.datadriveninvestor.com/predicting-tags-for-the-questions-i  
n-stack-overflow-29438367261e](https://medium.datadriveninvestor.com/predicting-tags-for-the-questions-in-stack-overflow-29438367261e)