

# SMAI Project Report - Team 40

## Multi-Factor Duplicate Question Detction in Stack Overflow

Ananya Sane      L Lakshmanan      Poorva Pisal      Sneha Raghava Raju

### **Abstract**

This report details our implementation of the paper titled “Multi-Factor Duplicate Question Detection in Stack Overflow” by Y. Zhang et al. The objective of the paper is to create and train a model that is able to take a new question from a user as an input, compute certain weighted measures and give a list of k questions that are most similar to the input question. The similarity between two questions is calculated by using four distinct parameters, that are used in computing the measures, that stem from the four components in a stack overflow question - Title, Description, Tags and Topics. The similarity score is calculated and then used as a metric for ranking and creating the output list. We implement the model using Python due to its extensive library support. We also train and test the model on a smaller dataset due to hardware limitations, however the model has the same architecture and hence, should scale well. We conclude by showing results and listing some limitations.