

MSc Project Report

“StagIt”

**Machine Learning based Real Time Multi-Tag Recommendation System
for Stack Overflow Questions in the domain of C# technology.**

Dissertation by:

Kalpani Udeshika Seneviratne

IIT ID:2018611

RGU ID: 1814878

Supervised by:

Mr. Cassim Farook

August 2021

A report submitted as part of the requirements for the degree of
MSc Big Data Analytics at Robert Gordon University, Aberdeen, Scotland

Abstract

Community Question Answering platforms have been trendy in recent times. These are web-based information systems that connect users with similar interests. Stack Overflow is one such forum popular among the software development community. The website uses tagging as a strategy to categorise question contents and connect questions with the most expert users in particular technological domains. Currently, the tagging feature operates as a manual process.

“StagIt” was implemented to eliminate the extra time and ambiguity from the developer activity by automating the tagging process. The application is designed as a Chrome plugin in the Stack Overflow website, and covers selected sub-domains of C# technology.

The proposed solution obtains the question title and body from users and predicts the tags on the fly. The system’s core leverages multi-label classification techniques and TF-IDF feature extraction and LinearSVC machine learning algorithm.

Keywords: Multi-label classification, LinearSVC, Stack Overflow, Machine Learning, TF-IDF