Bi-Weekly Report 1

12th October 2016

Team 1: Microsoft Capita Data Mining Project

Carlota Ortega Vega, Fasbeer Eskander and Bethany Graves

OVERVIEW

Over the past two weeks, we were assigned our teams and project. Initially, we received a brief about our project which enabled us to begin our research in data mining. We also had our first meeting with the client, who gave us a more detailed overview of our project, the client company and the end users of the data mining solution we will work on.

MEETINGS

4th October 2016

The first time we met was at the computer science lab where we were assigned our teams and the project. The members of our team introduced ourselves to each other and Microsoft personnel gave us the gist of the upcoming project. Later the TA Saheed met with us and advised us to brush up on our skills and learn the technologies and skills that might be required for the project at hand. We were later given the contact details of our client from Capita to set up a meeting and discuss the project in detail. Finally, we set up a Facebook group to be able to contact each other in the team.

12th October 2016

In our first meeting with our client, we learned about the company we are doing the project for: Capita. He explained to us that our project will be to create a data mining solution based on their school data management software, SIMS. Moreover, the company is currently working on combining each school's individual data center into a nation-wide data warehouse, that will allow better data analysis to be carried out. The software currently does not provide explicit data mining functions, which is what our team will work on. The client hopes we will be able to work on this project with the Institute of Education to organise a panel with teachers and other school staff in other to gather our requirements from the end users of the software.

TASKS COMPLETED

- Contacted and met client
- Discussed access to SIMS software and database with our client
- Created Facebook group to facilitate team communication and file sharing
- Started research on data mining techniques, software, APIs and programming languages

PROBLEMS

- Need to contact Institute of Education to collaborate with them, especially in the requirements gathering phase

FUTURE PLAN

- Continue research on data mining techniques, software, APIs and programming languages
- Contact Institute of Education to try to organise a panel with staff from schools to gather requirements regarding what type of data analysis they would like to have in the software
- Create project website
- Obtain access to SIMS software and data from client

INDIVIDUAL REPORTS

Carlota Ortega Vega

As the team leader, I was in charge of contacting the client and arranging the first meeting with him. I was the client liaison for both my team and team 2, as they are working on a related project. Furthermore, I started exploring data mining, since it is a completely new field for me, to better understand our project. As part of my research for the project, I have also been looking at SQL and Microsoft Azure, since they will be some of the main tools we will be using.

Fasbeer Eskander

Following the day we were assigned our team and projects, our team was mainly focused on researching about the technologies that would be needed for the project. After looking at some of the required skills, with advice from the TA Saheed, i brushed my my skills on Java, HTTP, and learned about C#, SQL and Databases. Furthermore I looked into a few online courses for Data Science specifically the ones available on Lynda.com to gain more knowledge on the sort of

project we would be doing. After the meeting with our client, I plan to further study on these for the project.

Bethany Graves

In the time running up to our client meetings, our team has been in a period of general research on data mining and its implementations. To this end, I have been researching both well known and innovative data mining techniques and approaches. This research consists of reading several important Data Science research publications, studying current and up-to-date data mining articles and perusing online courses. The topics of my research include but are not limited to: Feature selection, association analysis and the use of association rules to process large data sets, and Machine Learning.