

Team 4: Bi-Weekly Report 7

Project Title: PEACH Data Mining

Date: 9 February 2018

Gavin Shek, Saqib Jahangir, David Stepanovs

Overview

In the last couple of weeks, we produced an initial prototype of the data anonymiser with a newly designed UI, which we demonstrated to our client, and am also progressing with the visualisation aspect of the project. With the visualiser, we have set up the entire Elastic Stack in a Docker container and are currently working to run the existing core analytics system so that we can connect Elastic Stack with Spark.

Meetings

Client Meeting - 9th February

During this meeting, we demonstrated the data anonymiser prototype to our client. Navin was pleased with the result and suggested some ideas of improvement:

- Allow for date ranges
 - Especially for birthdates/events
- To add more supported fields
- To add step frequency and step interval options
 - Especially for ages/dates etc.
- Add NiFi component to allow automatic upload to the core analytics system

We aim to complete implementation for this during the next iteration of development.

Finally, Navin asked to have a Windows executable of the tool produced so that a research group will be able to use it to anonymise real data which can then be used to test the core analytics system.

Numerous Meetings During and Outside of Lab Sessions

We have had various meetings where we have planned and delegated our tasks. We have decided to use the GitLab board to keep track of issues, features to be developed and tasks to be done.

Completed Tasks

- Dockerised the entire Elastic Stack
- Created a new UI for the anonymisation tool

- Produced a Proof of Concept (PoC) of the tool using test data

Problems to be Solved

- Deploying the existing core analytics system both locally and on Azure
- Connect the existing system with the visualisation tool

Plan

- Add more features to the data anonymisation tool
- Connect the Elastic Stack with the core analytics system
- Produce a PoC of the entire analytics system using test data
- Deploy the existing core analytics system on Azure
- Carry out stress tests of the system on Azure

Individual Section

Gavin Shek

During these past 2 weeks, I set up Elastic Stack within a Docker container and have been working to get a PoC of the visualiser using test data. I have also found several connectors to connect the container with the existing core analytics and am currently trying them out. Furthermore, I have been setting up the GitLab board to keep track of everything the team has to do.

Saqib Jahangir

In the last 2 weeks, I completed running the existing Core Analytics system locally and I configured the properties of Ni-Fi processors that control data-flow. I spoke to last year's team for guidance to get the system running because of the ambiguities in the instructions provided.

David Stepanovs

During the past 2 weeks, I have worked on designing the UI of the data anonymisation tool. I have also refactored the last team's code and added additional features. Furthermore I worked on the data generator part.