Team 4: Bi-Weekly Report 5

Project Title: PEACH Data Mining   Date: 15 December 2017

Gavin Shek, Saqib Jahangir, David Stepanovs

Overview

Over the last 2 weeks, we have created a simple data anonymization prototype using the open source library ARX. We have decided to create this prototype with a simple user interface because the full ARX user interface is quite complex but allows for a lot of functionality. We are aiming to create a tool suitable for the users and so, have begun to use the open source library to create our tool.

Meetings

Numerous Meetings During and Outside of Lab Sessions

We have worked on a Java program which uses the ARX data anonymization library. Although the data generator tool has already been created by the previous team, we decided to familiarise ourselves with the ARX first by creating this prototype before integrating the existing data generator tool into our solution.

Completed Tasks

- Created a simple Java demo using ARX library

Problems to be Solved

- How to generate more data from a small set?
- Understand concepts like
  - k-anonymity, ℓ-diversity, t-closeness and δ-presence
  - (ε, δ)-differential privacy

Plan

- Design GUI for our app
  - Improve the usability
    - Add option to select which columns have to be anonymized
- Get a sample dataset which we can test our program with
- Integrate the original data generator tool to our project
- Integrate machine learning and related concepts to our tool

- Begin the integration of Kibana and Elastic Stack to the data analytics system

Individual Section

Gavin Shek

I have contributed to the development of the initial prototype using the ARX library, and tested our tool on fake data.

Saqib Jahangir

I have found multiple fake datasets which we used to test our data anonymization tool.

David Stepanovs

I have worked on the prototype and looked at how the previous team used the ARX library.