Team 4: Bi-Weekly Report 4

Project Title: PEACH Data Mining Date: 1 December 2017

Gavin Shek, Saqib Jahangir, David Stepanovs

Overview

Over the last 2 weeks we have been looking at the ARX library in further depth. We gained an understanding of how the library works and planned what we were able to incorporate into the project. We have decided to specialise in the ARX library because we had decided that this was the most suitable existing solution for our project. It is both open source and more effective than other solutions we had researched. Furthermore, the existing data generator is based on the ARX library and that is why we have decided to research this software in further detail.

Meetings

Numerous Meetings During and Outside of Lab Sessions

We have researched into the statistical concepts behind ARX and gained an understanding of the code. We also began gaining understanding of the structure of the code and coordinated research into different concepts including machine learning and genetic algorithms.

Completed Tasks

- Researched the ARX data anonymization library
- Looked at the code of the last year's team data anonymization tool
- Looked for alternative data anonymization tools which use
    - Genetic algorithms
    - Machine learning
    - Other new technologies

Problems to be Solved

- Write simple prototype using ARX
    - Input csv file with confidential information and output an anonymized csv file
- Understand the mathematical models used for data anonymization

Plan

- Program a simple data anonymization tool using the ARX library
- Plan how to embed advanced concepts into the tool

Individual Section

Gavin Shek

In these last two weeks, I have researched into machine learning, and related topics. I then planned how to embed this into our data anonymization tool.

Saqib Jahangir

I have looked at the code of the last year's data anonymization tool.

David Stepanovs

In the last two weeks, I have researched into the ARX data anonymization library.