

Detecting misuse in processes and data requests to Health Information System based on requests logs

Urmo Lihten

December 11, 2019

Problem Statement

Supervisor: Jaan Priisalu Co-Supervisor: Priit Raspel

1 Motivation

Health Information System in Estonia holds different electronic health records for almost every person in Estonia. Also for some foreigners who have visited a doctor or a emergency medical care facility for various reasons in Estonia. Every healthcare facility in Estonia has to send their patient medical records electronically to central Health Information System, so that patients can view these documents electronically through a website called Patient Portal and other healthcare facilities and doctors can view their patients previous documents and data through their software. This provides visibility between patient and doctor communicatoin and other healthcare facilities benefit from it by seeing previous data and documents when patients goes to other healthcare facility.

This provides doctors and nurses the ability look into patsients private data, when patient turns to them with a problem. This also gives the possibility to browse peoples data even when they arent treating the patient or the patient has never turned to them in any way. It is a security threat which usually is regulated with in-house rules and contracts, when a staff member of the healthcare facility has the right to access persons medical data. But

this is usually only regulated on paper and its difficult to check when the medical staff really has the right to check persons information. The right is given when a patient has turned to that healthcare facility and their medical staff but any other time doctors and nurses dont have the right unless it is releated to their previous relationship and/or encounter associated with a medical procesdure.

This gives motivation to examine the usage of Health Information System and detect anomalies and data misuses by the healthcare facilities and their staff using machine learning and process mining using feature selection and other methods depending on the information what kind of usages are wrong and what to look for in the request logs.

2 Research problem

Health Information System gives doctors and nurses alot of access to private data and it is hard to determine, if patient has really turned to them for medical help or not in a direct way or through another way or procedures what are ordered by the doctor. In another way, when in an emergency and patient is un-cooperative or in such state, unable to communicate and identity has to be confirmed without an consent to save patients life in a critical situation then this poses a possible security threat. Cause rights cannot be explicitly given to the doctor to view certain patients data in a critical situation. This allows medical staff to open any persons medical history and view it at any given time wheter the person has any medical relationship to that medical staff.

When a persons private data such as medical information is viewed and used, then there has to be a reason. Even if it wrongly done but is still explainable (wrong id-code submission by accident due to similarities of the ID-codes or the other person has given wrong patient ID-code by accident).

To solve this problem is to detect health records data misuse as early as possible with the help of machine learning and input from given processes about when each data request should be made and what should have been next. After that notifying the proper medical facility and if the misuse is very serious, then to the proper authorities.

3 Scope

Basis of this is the national X-road system logs with the requests and data from medical facilities to the Health Information System in Estonia. This contains clients and service information and also hl7 standard type documents and request about different types of medical documents of any person who have been inserted to the system through different medical assessments and documentation.

4 Literature review and novelty to the other studies

Most of the case studies and previous work are surrounding one or two same group hospital but not through an entire nation. In Estonia all medical institutions are required to send their patients data electronically to the Health Information System. Some works and studies only include a very short time period where machine learning is used and this giving a quite small scope and smaller possibilities for the machine learning classifier to learn and find anomalies within the processes and logs. Since most of the countries have not adopted x-road services in healthcare, this gives another angle to view and study about data access and usage. Also the time period for these x-road logs stored is longer than a few months providing the possibility to let the classifier learn from much more data with given processes on a national scale through out medical institutions. Bibliography to firstly viewed and researched problems on similar issues is added.