Healthcare AI startup

Project Description

Development of a tool for detection of abnormal behavior (for instance, VIP snooping, looking of disease history of colleagues, record views by terminated employees, altering patient diagnosis and so on) of hospital employee (doctors, nurses, etc.) who use working stations with all patients' data based on log activity. Creation of the anomaly detection tool to be easily integrated into the main system.

CUSTOMER

NDA

INDUSTRY

Healthcare

Anomaly detection

TYPE

Data science

TECHNOLOGY

Python

PySpark

parqeut

BigQuery

Amazon Redshift

libfm & libffm

scikit-learn

plotly

matplotlib

Challenges

- Developing a high-load and effective system architecture for writing, reading and processing logs data;
- Understanding the meaning of the data fields and their importance;
- Preparing a list of all typical scenarios of abnormal behavior able to be retrieved from the data;
- Implementing anomaly detection;
- Distracting data from the storage (written in text files).

Solutions

- Application of Amazon Redshift for storing fast increasing amounts of data that provided effective access to the data.
- Preparation of the list of different types of suspicious users' activity that should be detected either right after it appeared or in the scheduled mode.
- Implementation of various approaches based on time series anomaly detection, classification algorithms such as isolation forest and one-class-SVM, clustering and dimension reduction techniques.
- Design of two unique approaches based on factorization machines. For hypothesis testing and model building, some of the data were saved both in parquet files and in BigQuery tables for the fast processing in Apache Spark and other libraries.
- Proposition and development of the approach based on clustering algorithms and time series statistical methods for searching patterns of abnormal users' behavior.

Summary

We have developed a system for monitoring the use of data from hospital patients, allowing identification of suspicious actions of users with a very small amount of false negatives; notifying security officers about any abnormal action right after it was done; investigation and analysis of the users' activity visualized as dashboards; searching for suspicious patterns in the behavior of system users for unknown use-cases.