Menshikh Ivan

menshikh.iv@gmail.com | menshikh-iv

Education

IMCS UrFU
(Institute of Mathematics and Computer Science in Ural Federal University)
Ekaterinburg, Russia

B.S. in Fundamental Informatic and IT
2012 - 2016

IMCS UrFU
(Institute of Mathematics and Computer Science in Ural Federal University)
Ekaterinburg, Russia

M.S. in Computer Science
2016 - PRESENT

Skills

Programming Python, SQL, C++, Java
Big Data Spark, Cassandra, Hbase, Hadoop
DevOps Ansible, Docker
Web Flask, HTML, CSS, Js

Experience

SkyDNS
Ekaterinburg, Russia

Data Scientist & Software Engineer
Jul. 2015 - PRESENT

- **Implement malicious domain activity scoring**
  In solution, I construct the bipartite graph (users and domains) and calculate several suspicious and non-suspicious score for a graph (an iterative process, which transfers score from domains to users and back) and some other features. After that, I used machine learning technics for union this scores and predict the final suspicious score. On the basis of this development, I wrote my graduation work.
- **Implement web page classification**
  Using a crawler to collect web pages from the internet and need to have a category for every website/web page for content filtering. I used multiple semantic representations for content from a web page (including LDA, LSI, and Doc2Vec from gensim) for the classification task.
- **Implement similar page search**
  I used MinHash with buckets for duplicate search (for example, detect parked sites) and fuzzy search based on LSI with Annoy for detecting sites with very similar content.

Extracurricular Activity

Open source contributions

- **Gensim PR#782** - Add functionality for distributed LDA to train this in global network.
- **Gensim PR#1154** - Small fix that significantly improves a performance of applying LDA model in default case for large models.

Own project

- **image2pic** - Semantical search based on query text or query pictures. We used bigARTM as a multimodal topic model for all discrete distributions (comment, tag, and labels from inception) and Inception as label generator for pictures (with box-cox transformation).