

Alexandr Koryachko

akoryachko@gmail.com | 06.214.666.06 | koryachko.com

SKILLS

PROGRAMMING

Languages:

Python • PySpark • SQL • Matlab

Packages:

Pandas • Polars • Scikit-learn • PyTorch
numpy • NLTK • spaCy • mllib • Keras

Tools:

Git • Azure Databricks • AWS • \LaTeX

DATA SCIENCE

Big Data • Statistics • Machine Learning
Data Visualization • Experiment Design

LINKS

LinkedIn:// [akoryachko](#)

Github:// [akoryachko](#)

GoogleScholar:// [bISiDH4AAAAJ](#)

EDUCATION

NORTH CAROLINA STATE UNIVERSITY

PHD IN ELECTRICAL ENGINEERING

Aug 2018 | Raleigh, NC

Minor in Mathematics

GPA: 4.0 / 4.0

NATIONAL RESEARCH UNIVERSITY OF ELECTRONIC TECHNOLOGY

MS IN COMPUTER SCIENCE

Jul 2009 | Moscow, Russia

Conc. in Software Engineering

GPA: 4.0 / 4.0

BS IN COMPUTER SCIENCE

Jun 2007 | Moscow, Russia

Conc. in Information Security

GPA: 3.92 / 4.0

COURSEWORK

Pattern Recognition

Detection and Estimation

Neural Networks

Mathematical Modeling

Uncertainty Quantification

Applied Bayesian Analysis

Random Processes

Graphs and Graphical Models

Computer Vision

Information Theory

Digital Signal Processing

Nonlinear Programming

Multivariable Control Systems

Numerical Analysis

Modeling of Biological Systems

Technical Communications

EXPERIENCE

XOMNIA | DATA SCIENTIST

Apr 2023 - Present | Amsterdam, Netherlands

Consulting at Rabobank:

- Initiated and took a leading role in the periodicity cold start project. The project aims to predict a periodic sequence from a single transaction.
- Advised a team of young professionals in developing a GDPR-sensitive transactions detector using the Azure platform tools.
- Implemented a statistical test for performance stability of the transaction periodicity detection model with PySpark in Databricks.

VERICAST | STAFF DATA SCIENTIST I

Feb 2022 - Sep 2022 | Morrisville, NC, USA

Developed automated feature engineering functionality for big data applications in a scalable machine learning execution engine on AWS. The corresponding Python library was projected to save up to 90% time a Data Scientist spends writing custom SQL queries for each new client's dataset.

VERICAST | DATA SCIENTIST II

Mar 2021 - Feb 2022 | Morrisville, NC, USA

Prototyped tools for a privacy centered digital ad targeting solution that relies on webpage content as opposed to user tracking for placing clients' ads:

- Built machine learning pipelines that analyze 300 million crawled web pages to select digital ad placements that are the most relevant to clients' desired content from 10 billion options daily.
- Trained an embedding model for 3,000 interest categories based on 200 million classified web URLs with PySpark Machine Learning library. The model suggests the most similar categories to clients' keywords. The implementation reduced ad campaign setup time and increased the number of relevant categories to target.
- Built a classification model for identifying web pages with brand sensitive content. The model reduced ad serving on websites with unsafe content by 80% and improved clients' satisfaction with reporting.

VERICAST | DATA SCIENTIST I

Dec 2018 - Mar 2021 | Morrisville, NC, USA

Designed Machine Learning algorithms for Big Data processing

- Developed machine learning models for quantifying users' interests in a variety of topics by analyzing billions of daily user URL visitations with PySpark Machine Learning toolkit.
- Engineered tools for identifying store competitors from tens of millions of possible options by matching descriptions with spaCy Python library.
- Designed and tuned a random forest model for inventory quality prediction for digital ad auctions using Scikit-Learn. The resulting model improved the average ad viewability scores by 5%.

RESEARCH

ENBISYS LAB | RESEARCH ASSISTANT

Aug 2012 - Jul 2018 | Raleigh, NC, USA

Developed computational approaches for large-scale genomics data analysis

- Reduced the search space for iron deficiency induced regulators from a set of almost 3,000 differentially expressed genes to a testable subset of 7 candidates by applying unsupervised machine learning algorithms.
- Mined data from diverse experimental datasets to form prior probability distributions for the model parameters.
- Quantified uncertainty associated with the model fit to produce testable predictions which were then validated with additional experiments.