ARTEM VOLGIN

Data Scientist

artvolgin.github.io
art.volgin@gmail.com
+447835950253
github.com/artvolgin
Manchester, UK
in/artvolgin

SUMMARY

- Data Scientist with diverse experience working with large-scale behavioural and financial data in marketing research, non-profit sector, and academia.
- Proficient in presenting results to a broad audience and working with various stakeholders.
- A passionate participant in data science competitions.

SKILLS

Toolkit: Python (pandas, numpy, scikit-learn, xgboost, seaborn, beautifulsoup, pytorch, huggingface, spacy, nltk, statsmodels, networkx), R (dplyr, ggplot2, brms, igraph), SQL, MongoDB, Spark, Docker, AWS, Tableau, Git

Expertise: Machine Learning, NLP, Causal Inference, Bayesian Statistics, Spatial and Network Analysis

Languages: English - fluent, Russian - native

EXPERIENCE

9/2021 – 6/2024 **Graduate Researcher** **University of Manchester, UK**

- Implemented a pipeline using Large Language Models (LLMs) for named entity recognition and relation extraction from documents related to corruption incidents.
- Analysed the influence of private companies on educational institutions using advanced network simulations and hierarchical Bayesian models in combination with a large company database.
- Examined the effect of the war on migration patterns using unique time-series web search data.
- Led seminars on Introduction to Statistics and Network Analysis for over 100 students.
- Presented my research at multiple conferences and published several papers.

2/2022 – 9/2022 **Graduate Researcher** **University of Oxford, UK**

- Scrapped, matched, and deduplicated genealogical profiles and financial records from different sources.
- Created a pipeline for classifying ethnicity and gender based on individuals' names.
- Extracted educational and career trajectories from text with pre-trained deep learning NLP models.

5/2019 – 5/2021 **Data Scientist** **The World Bank, Russia**

- Developed a novel approach for calculating the value of skills acquired from universities, utilizing supervised machine learning methods and named entity deduplication on extensive resume datasets.
- Evaluated the accessibility of extracurricular education organisations by applying methods from spatial statistics based on linked data from different administrative sources.
- Constructed and maintained a database with financial indicators of educational organizations.
- Estimated financial returns to education using causal inference techniques based on the panel data.
- Delivered presentations of my work to various internal and external non-technical stakeholders.

1/2016 – 5/2019 **Senior Data Analyst** **Russian Public Opinion Research Center, Russia**

- Contributed to over 20 projects for marketing companies and nonprofit organizations.
- Analysed survey data with advanced statistical methods.
- Worked directly with clients to design and implement market research that meets their objectives.
- Organised and conducted training for interns and junior colleagues.

EDUCATION

9/2021 – 6/2024 **PhD** Social Statistics **University of Manchester, UK**

9/2018 – 6/2020 **MS** Applied Statistics **Higher School of Economics, Russia**

9/2011 – 6/2015 **BA** Social Sciences **Russian State University for the Humanities, Russia**

DATA SCIENCE COMPETITIONS

2023 **Unsupervised Wisdom by CDC** – 1st place 25,000$
Applied LLMs in combination with an SVM classifier to label a large dataset of medical texts.

2020 **COVID-19 Symptom Data Challenge by Facebook** – 2nd place 30,000$
Analysed the impact of COVID-19 policy with a causal time-series model on a 10 million Facebook survey.

2020 **Unlocking Climate Solutions by CDP** – 2nd place 25,000$
Created a method for analysis of environmental reports and surveys of managers with data envelopment analysis, association rules mining, and topic modelling.

2019 **Environmental Insights Explorer by Google** – 1st place 10,000$
Developed a method to estimate the region's annual emissions factor using remote sensing data emissions, weather conditions, and OpenStreetMap information using spatial-temporal models.