ARTEM VOLGIN

Data Scientist

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

SUMMARY

Data Scientist with experience working with behavioural and financial data in marketing, non-profit sector, and academia. Proficient in translating complex questions into data-driven insights and presenting them to a non-technical audience. An active and passionate participant in data science competitions, regularly achieving top positions.

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

Toolkit: Python (pandas, numpy, sklearn, xgboost, seaborn, beautifulsoup, selenium, pytorch, transformers, spacy, nltk), R (tidyverse, data.table, sf, brms, igraph), SQL, Spark, MongoDB, Docker, AWS, Tableau

EXPERIENCE

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

- Implemented a pipeline for named entity recognition and relation extraction from official descriptions of corruption incidents using Large Language Models (LLMs).
- Used hierarchical Bayesian models and a large business ownership database to estimate the influence of private companies on educational organisations.
- Performed large-scale stochastic network simulations using AWS cloud platform.
- Scrapped granular web search data to examine the effect of the war on migration-related queries.
- Led seminars on Introduction to Statistics and Network Analysis for over 100 students.

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

- Scrapped, matched, and deduplicated over 1 million genealogical profiles and financial records of the British elite from 5 online sources, obtaining a unique historical dataset.
- Used pre-trained deep learning NLP models to extract educational and career trajectories from biographies and classify ethnicity and gender based on individuals' names.

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

- Employed supervised ML methods and named entity deduplication on extensive resume and vacancies datasets with over 25 million observations for estimating skills provided by universities.
- Utilised spatial statistics and large administrative data to evaluate the accessibility of training facilities.
- Applied causal inference methods to estimate financial returns to education based on the panel data.
- Constructed and maintained a MongoDB database with financial indicators of educational organisations.
- Delivered presentations and policy recommendations to various internal and external stakeholders with working papers downloaded over 3,000 times.

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

- Contributed to over 30 research projects for marketing companies and nonprofit organisations.
- Conducted analysis of survey data using advanced statistical techniques.
- Worked directly with clients to design and implement quantitative market research.
- Organised and delivered 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

10/2023 **1st place** – Unsupervised Wisdom by CDC – 25,000$
Applied LLMs in combination with an SVM classifier to label a large dataset of medical texts.

11/2022 **4th place** – Big Data Derby 2022 by NYRA – 10,000$
Extracted racing strategies from tracking data using clustering methods and assessed their effectiveness.

12/2020 **2nd place** – COVID-19 Symptom Data Challenge by Facebook – 30,000$
Analysed the impact of COVID-19 policy with a causal time-series model on a 10 million Facebook survey.

11/2020 **2nd place** – Unlocking Climate Solutions by CDP – 25,000$
Analysed environmental reports using association rules mining and topic modelling.

3/2019 **1st place** – Environmental Insights Explorer by Google – 10,000$
Estimated emission factor using remote sensing data, OpenStreetMap, and spatial-temporal models.