Szymon Socha

https://szymonsocha.github.io/

✉ upon request | 📱 upon request
🔗 szymon-socha | ⬛ szymonsocha

EXPERIENCE

Data Scientist | *Python, PySpark, Apache Spark, Machine Learning*   Feb 2024 – present
TEAM International / PwC Czech Republic   *Remote, PL*
- Developing, improving, and optimizing the AML system code in Python and PySpark
- Optimizing transaction monitoring (TM) scenarios by applying statistical methods and machine learning algorithms
- Delivering analysis results and insights in the form of visualizations and reports for clients in the banking sector
- Member of the Financial Crime Technology team

Data Scientist | *Python, Machine Learning, NLP, Time Series Forecasting, LLM*   Mar 2023 – Dec 2023
WeSub   *Warsaw (Remote), PL*
- Conducting R&D to develop new models for product categorization and value prediction within a circular economy framework, using machine learning and artificial intelligence algorithms
- Developing a range of **Time Series Forecasting** models by leveraging Natural Language Processing (NLP) and **transfer learning** features
- Researching cutting-edge techniques for employing and **fine-tuning of Large Language Models** (LLMs)
- Creating **NLP** models for text generation and classification
- The National Centre for Research and Development (NCBR) project

Data Scientist (Junior) | *R, SQL, BigQuery, ETL*   Jan 2022 – Apr 2023
Flying Wild Hog   *Warsaw (Remote), PL*
- Analyzing video games telemetry data using **R** and **SQL**
- Developing dashboards and reports in **RMarkdown** and **Shiny**
- Developing **ETL** processes with **BigQuery** and **MongoDB**
- Collaborating with UX Research team members in combining quantitative and qualitative research findings
- Working closely with different game development teams and editorial teams around the world in order to understand their requirements and questions related to a project

Data Science Assistant | *Python, R, Econometric Modelling*   Sep 2021 – Nov 2021
Publicis Groupe   *Warsaw, PL*
- Developing tool for econometric modeling project
- Data cleaning and processing using **Python** and **R** and building custom R tools for econometric analysis

EDUCATION

Master of Science in Data Science and Business Analytics
University of Warsaw
GPA: 5/5, **Awards**: Dean's List (Top 7%)
Thesis: *Should You Care about Tools in Fake News Detection? A Comparative Study of Cutting-Edge Embedding Methods and Classification Algorithms*
Courses: Machine Learning I & II, Unsupervised Learning, Advanced Econometrics, Applied Finance, Statistics, Python, Advanced Programming in R, Big Data Analytics, Algorithms for Data Science, Bayesian Time Series Forecasting

Bachelor of Science in Computer Science and Econometrics
University of Warsaw
Thesis: *The Impact of the COVID-19 Pandemic on Music Streaming Consumption*
Courses: Machine Learning, Econometrics, Linear Algebra, Statistics, Time Series, Probability Calculus, Finance, Python, R, SQL, C++

SELECTED PROJECTS

Option pricing Rcpp package | *C++, R, Rcpp, Rmd*
- Path-dependent European style up-and-out call option pricing with Monte Carlo and Rcpp package

TECHNICAL SKILLS

Languages: Python, R, SQL (BigQuery, KQL)
Tools: Git (GitLab, Bitbucket), FastAPI, Google Cloud Platform, PyCharm, RStudio, Ubuntu
Libraries: PyTorch, PyTorch Lightning, HuggingFace, PySpark, sklearn, pandas, NumPy, tidyverse, dplyr, ggplot2
Expertise: machine learning, natural language processing (NLP), time series forecasting, econometric data modelling, clusterization, dimension reduction, statistics, hypothesis testing