

# Szymon Socha

<https://szymonsocha.github.io/>

✉ *upon request* | 📄 *upon request*  
in szymon-socha | 🌐 szymonsocha

## EDUCATION

### University of Warsaw

*Master of Science in Data Science and Business Analytics*

Warsaw, PL

2021 – 2023

**GPA:** 5/5, **Awards:** Dean's List 2022 (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, Machine Learning II, Unsupervised Learning, Advanced Econometrics, Applied Finance, Statistics, Python, Advanced Programming in R, Big Data Analytics, Algorithms for Data Science, Bayesian Time Series Forecasting

### University of Warsaw

*Bachelor of Science in Computer Science and Econometrics*

Warsaw, PL

2018 – 2021

**Thesis:** Investigating the impact of the COVID-19 pandemic on music streaming consumption, utilizing Google Trends activity changes and Spotify statistics. An econometric model is constructed and evaluated using R and Python

**Courses:** Machine Learning, Econometrics, Linear Algebra, Statistics, Time Series, Probability Calculus, Finance, Python, R, SQL, C++

## EXPERIENCE

### Data Scientist

*WeSub*

Apr 2023 – Present

Warsaw, PL

### Data Scientist (Junior)

*Flying Wild Hog*

Jan 2022 – Apr 2023

Warsaw, 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

*Publicis Groupe*

Sep 2021 – Nov 2021

Warsaw, PL

- Developing tool for econometric modeling project
- Data cleaning and processing using **Python** and **R** and building custom R tools for econometric analysis

### QA Tester

*Testronic*

Jan 2018 – Oct 2018

Warsaw, PL

- Detecting and reporting bugs, software flaws and design issues (**JIRA**)

## SELECTED PROJECTS

### Neural Network Performance Comparison | *Python, Neural Network, LightGBM, XGBoost, Random Forest*

- Comparison of neural network performance with other ML algorithms for two different types of data (image classification and regression)

### Traffic Prediction | *Python, sklearn, Random Forest, Linear Regression, KNN, SVM, Rmd*

- Traffic prediction using multiple ML algorithms. Data manipulation and feature engineering

### 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), Google Cloud Platform, PyCharm, RStudio, Ubuntu

**Libraries:** sklearn, pandas, NumPy, tidyverse, dplyr, ggplot2

**Expertise:** machine learning, natural language processing (NLP), econometric data modelling, clusterization, dimension reduction, statistics, hypothesis testing

I agree to the processing of personal data provided in this document for realising the recruitment process pursuant to the Personal Data Protection Act of 10 May 2018 (Journal of Laws 2018, item 1000) and in agreement with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).