Szymon Socha

szymon.socha13@gmail.com | +48603820390| linkedIn/szymon-socha | github/szymonsocha | szymonsocha.github.io

SUMMARY

Data Scientist/Analyst at Flying Wild Hog. Data Science and Business Analytics final year student at the University of Warsaw. Uses statistical modelling and machine learning algorithms to finding new patterns and relationships in data to drive business value. Passionate about using data to solve real-world problems. Avid traveler.

EDUCATION

MSc. Data Science and Business Analytics — Warsaw, PL | 2021 - present
University of Warsaw
Coursework: 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

BSc. Computer Science and Econometrics — Warsaw, PL | 2018 - 2021
University of Warsaw — **GPA:** 4.5/5
Thesis: "The influence of COVID-19 pandemic on the music streaming consumption." An econometric OLS model constructed and tested with R programming language. Python programming language used for data manipulation
Coursework: Econometrics, Linear Algebra, Statistics, Time Series, Probability Calculus, Finance, SQL, R, C++

WORK EXPERIENCE

FLYING WILD HOG | Data Scientist/Analyst (Junior) — Warsaw, PL | Jan 2022 – present

PUBLICIS GROUPE | Data Science Assistant — Warsaw, PL | Sep 2021 - Nov 2021

- Econometric modelling project.
- Data cleaning and processing with the use of **Python** and **R**.
- Building custom R tools for econometric analysis.

TESTRONIC | QA Tester — Warsaw, PL | Feb 2018 – Oct 2018 and Jul 2020 – Oct 2020

- Detect and report bugs (in English), software flaws and design issues on a daily basis (**JIRA**).

PROJECTS

TRAFFIC PREDICTION — Python, sklearn, Random Forrest, Linear Regression, KNN, SVM, Rmd
Traffic prediction using multiple machine learning algorithms (Linear, Lasso, Ridge, and Elastic Net Regression, Random Forrest, KNN regressorm, SVM). Data manipulation and feature engineering.

DRUG USE CLASSIFICATION — Python, KNN, Logistic Regression, SVC, Rmd
Drug use prediction using multiple machine learning algorithms (Random Forrest, k-Nearest Neighbors Classifier, Logistic Regression, Support Vector Classifier). Data manipulation and feature engineering. Balancing data with SMOTE+Tomek. Cross Validation used for parameter tuning.

STUDENT DROPOUT PREDICTION — R, Logit, Probit, Hypothesis Testing
Prediction of student dropout using Logit and Probit models. Hypothesis testing to find the best model. Comparing Logit and Probit models performance. Interpreting the results and making conclusions.

SKILLS

Tech stack: R (tidyverse, dplyr, ggplot2, Rmd, Shiny), Python (sklearn, pandas, numpy), SQL (BigQuery, KQL), Git
Expertise: Machine Learning and Econometric data modelling, clusterization, dimension reduction, statistics, hypothesis testing
Soft skills: Problem-solving, Stress management, Decision-making, Time management, Communication, Attention to detail