SILVIO SOPIC

Data Scientist

DATA SCIENCE | DATA ENGINEERING | DATA VISUALIZATION | MARKETING

/LinkedIN
/Github
/email

Passionate about data-driven decision-making, I hold an MBA and dual MSc degrees (Business Economics and Information Management) from KU Leuven. Recognized for enhancing mass house price prediction models in a competitive thesis competition, I am skilled in R, SAS, Python, Bash, SQL, and data science libraries, excelling in conveying complex insights effectively. During a dedicated 6-month sabbatical, I furthered my expertise and achieved prestigious certifications, including those from IBM, in Data Analysis, Data Warehousing, Data Science, Machine Learning, and People and Soft skills.

EXPERIENCE

08/2022 - 01/2023 **Young Graduate Program | Functional Analyst**

KBC

- Completed first cycle of the straight through processing by implementing a python script to generate SQL queries to be used to scan for nulls in Teradata
- Undertook tasks in many individual parts of the ETL process as a Functional Analyst.
- Coordinated effectively with the security department which ensured proper distribution of technical privileges in the fine grained access system
- Updated internal codebooks using knowledge gained using SAP PowerDesigner and Teradata
- Attended the rituals of the scaled agile framework (SAFE)

10/2021 - 04/2022 **Functional Analyst**

Accenture

- ETL output comparison between Snowflake and SAP HANA using Excel and conditional formatting
- Internal ART: AI for Financial services member
- Developing prediction models for financial services (e.g., stock exchange, risk)
  - Responsible for analysis in the field of Credit risk modeling
- Business Analyst in PowerBI trainings (60h), Azure fundamentals, and Azure data fundamentals

EDUCATION

09/2018 – 09-2021 **MsC Information Management | Msc Business Economics**

KU Leuven (Times Higher Education Ranking: 42nd)

- MSc in Information Management
  - Successfully completed courses related to Data Science, Data Engineering, Data Analysis, and Software Development.
  - Thesis: "The impact of POI data on price prediction models."
  - Nominated for LIRIS thesis competition in the category of "POI in real estate price prediction"
  - Feature engineering using distance calculation in Euclidean space.
  - Developed dashboards in Tableau (3).
  - Implemented prediction models using random forest regressor with k-fold optimization.
- MSc in Business Economics
  - Completed an experiment replication with technical analysis using ANOVA.
  - Conducted mediation and moderation analysis.
  - Performed survey analysis.
  - Obtained Google Cloud certification during the program.

04/2015 – 06/2018 **General MBA**

Zagreb School of Economics and Management (QS Global MBA Ranking: Top 200)

- Equipped with a holistic business understanding and strategic thinking mindset.
- ERASMUS student exchange at KU Leuven (2 semesters)

LANGUAGES

Croatian C2 Proficient User | Mother tongue

English C2 Proficient User

German B1 Independent User

ACCOMPLISHMENTS

First Runner Up
LIRIS Best thesis of Information Management

A volunteer experience
International meeting center KU Leuven.

INTERESTS

Machine Learning
Data Science
Data Visualization
Data Engineering

SKILL SUMMARY

Programming languages

- Python ●●●●●
- SQL ●●●●○
- Bash ●●●○○
- SAS ●●●○○
- Git ●●●●○
- R ●●●●○

ML/DL libraries

- Scikit learn ●●●●●
- Tensorflow (Keras) ●●●○○
- Pyspark ●●●●○
- Openai ●●●●○
- XGBoost ●●●●○

Dataviz libraries

- Seaborn ●●●●●
- Matplotlib ●●●●○
- Folium ●●●●○
- Yellowbrick ●●●●○
- ggplot ●●●●○

Miscellaneous Libraries

- Pandas ●●●●●
- Numpy ●●●●○
- Imblearn ●●●●○
- Re ●●●○○
- Lifetimes ●●●●○

Dataviz software

- Tableau ●●●●●
- PowerBI ●●●●○
- Cognos ●●●○○

Softskills

- Prompt engineering ●●●●●
- Presenting ●●●●○
- Problem Solving ●●●●●
- Collaboration ●●●●○
- Brainstorming ●●●●●

Databases (libraries)

- Pyspark ●●●●●
- IBM db2 ●●●●○
- MySQL ●●●●○
- PostgreSQL ●●●●○
- Teradata ●●●●○
- Snowflake ●●●○○

Design and Documentation

- EER ●●●●●
- UML ●●●●○
- Latex ●●○○○

CODE SAMPLE

The impact of POI data on price prediction models

- Features the usage of distance matrices for feature engineering and cross-validation for hyperparameter tuning. Explores the usage of tree based models.

To Oversample or Undersample?

- Showcases the imb-learn package to increase prediction quality when dealing with grossly imbalalnced data. ML aglorithm wise, focuses on comparion various predictors and their stacking options.

Unravelling the Essence of Effective Market Segmentation

- Explores clustering through the initial steps of the marketing process. Aims to create a summary of the techniques and knowledge necessary to optimize customer intelligence.