

Jan Benisek

SENIOR DATA ENGINEER · MACHINE LEARNING ENGINEER · PYTHON DEVELOPER

7+ YEARS OF EXPERIENCE · MS AZURE CERTIFIED

EU citizen, currently relocating to Switzerland

☎ (+32) 497-406-353 | ✉ benisekj@gmail.com | 📷 JanBenisek | 📱 janbenisek

Experience

Python Predictions, by Tobania

Brussels, Belgium

SENIOR DATA ENGINEER CONSULTANT

06/2019 - today | 2y

• Internal roles:

- **Expanded** Data Engineering activities within our company by **hiring** data engineers, setting up **strategy**, growth plan and promotion materials. Successfully drafted and **sold multiple medium-sized projects**, valued at ~170k.
- Responsible for **technical screening** of new Data Scientists. Technical on-boarding and coaching of new colleagues as well as thesis and internship **guidance**.
- **Product owner** and active maintainer of internal open-source Python library **Cobra**.
- Member of action group which proposed structured **career path and skills** development.

• MLOps framework implementation | Banking | 4m:

- Designed MLOps framework for data science **model life-cycle**, built on top of **KubeFlow** with custom logging and monitoring using **ElasticSearch** and Kibana.
- The solution allowed Data Scientists to deploy and monitor new models within **minutes, instead of days**, without manual work.

• Customer data platform development | Banking | 4m:

- Collected and implemented changes to the **ETL** in **Redshift** which creates customer analytical records for predictive modelling.
- Built **automated tests** for evaluation of the final table using EMR and PySpark to ensure quality.
- As a result, Data Scientists were able to deliver **predictive models within one sprint**, without the need to clean and prepare data.

• Financial reporting and visualization | Internal | 1m:

- Developed a financial dashboard by combining data from multiple APIs. The solution was built with **Flask** on **EC2** and s3, hosted with **Beanstalk** and Route 53. Front-end was designed with Dash, including user **authentication**.
- The product was **migrated** from Google Sheets with manual process, which allowed managers to have instant overview of the financial situation and **saved 1 hour every week** in administration.

• In-house system for personalized offers | Retail | 7m:

- **Reduced dependency** on external vendor by building an in-house system in **MS Azure** for **personalized offers**. The system included processing of millions of purchase records, aggregated analytical tables and predictive models.
- Ensured on-time delivery of recommendations to customers by **deploying predictive models** and data pipelines using **Data Factory**.
- Increased **speed of predictive model from hours to minutes** by vectorizing code and idiomatic use of Numpy and SciPy.

• Targeted TV advertising | Telco | 3m:

- Processed TBs of viewing data on **PySpark**, allowing Data Scientists build predictive models. The project resulted in **fast and automated** data delivery and increased revenue.
- Programmed **automated model comparison**, which helped Data Scientists to detect when a predictive model becomes obsolete and to replace it, reducing manual work.

• Best-fit mobile tariff recommendation | Telco | 3m:

- Redesigned data preparation and model building of tens of ensemble models for tariff recommendations using **PySpark**. The solution **reduced manual work from days to hours** and allowed for on-demand model retraining and deployment.
- Built data pipelines for analysis of **spatial data**, further enriching the recommendation models.

• Optimizing candidate screening | Banking | 1m:

- Optimized candidate screening and time-to-hire by **scraping tens of thousands** of CVs and CLs using Python. The HR department was able to detect candidates more likely to succeed based on key words and skills.

DATA SCIENTIST CONSULTANT

10/2017 - 05/2019 | 1y8m

• Predictive models for marketing squad | Telco | 8m:

- Built a model which detected customers likely to use a self-install of certain products. During lockdown, the project significantly **lowered costs** and **increased customer satisfaction**.
- Rebuilt a model to **detect** customers who are likely to **purchase a second SIM card**. The uplift of the model was increased by 13%.
- Built a model to identify potential **churners**, who were targeted in anti-churn campaigns. The model was able to find 8.7% more churners than the current approach in top decile.
- Predicted competitor's next steps by analyzing **open data**. The result of the analysis was the basis for a new strategy in high-risk regions.

• Predicting credit risk | Retail | 2m:

- Found **new business opportunities** by processing and analyzing credit risk data using R. The analysis convinced the client to adapt and expand their credit risk evaluation process.
- Refitted credit risk models to be more flexible, allowing the risk specialist to counter-offer the applicants **viable alternatives**. The new model offered higher customer **satisfaction** and increased number of approved credits without increasing the risk.

• Financial product recommendation | Banking | 8m:

- Implemented and evaluated multiple models to **estimate purchase** of financial products. Marketing campaign resulted in a 5.4% increase in number of purchase.
- Built two **recommender engines** to offer relevant financial products to customers. The models allowed for personalized marketing campaigns and improved the model accuracy by 11.3%.

Test-Achats

Brussels, Belgium

DATA SCIENTIST

09/2015 - 10/2017 | 2y1m

- **Co-founder and tech. leader** of Advanced Analytics activities in multiple departments and branches in other countries. Responsible for technical solution and methodology. Participation in **hiring** of new analysts.
- Designed **churn prediction** model for better customer retention. Implemented data-driven **segmentation** to improve targeting and developed **insights** into customer behaviour to increase loyalty.
- Close cooperation with stakeholders, higher management, DWH team, business analysts and reporting specialists.

BUSINESS INTELLIGENCE SPECIALIST

09/2014 - 09/2015 | 1y

- One-year internship. **Data analysis** and visualization to support business decisions.
- Responsible for technical solution of requests - SQL and MS Excel **development**. Increased productivity with newly developed tools.
- Cooperation with DWH team - monitoring of data quality, participating in adding new facts and dimensions.

Ceska Sportelna

Prague, Czech Republic

DATA ANALYST

12/2013 - 08/2014 | 9m

- Helped with design of new **reporting** system. Responsible for technical solutions. Fixed-term contract.
- Regular and ad-hoc reports and **analysis** for management and branches. Improved system for faster remuneration for salesmen.

Education

Azure Data Engineer Associate

MICROSOFT CERTIFICATIONS

10/2020 - 04/2021

- DP-200: Implementing an Azure Data Solution
- DP-201: Designing an Azure Data Solution

Statistics, Msc.

Leuven, Belgium

KU LEUVEN, FACULTY OF SCIENCE

02/2016 - 09/2019

- Two-years program, specialization in Business Statistics. Part-time. **Cum Laude**
- **Master thesis**: Software for automated diagnosis of ovarian cancer from ultrasound images.
 - In cooperation with UZ Leuven hospital, I proposed a U-Net model in Keras for automated lesion segmentation.

Skills

Languages

English: Full proficiency (IELTS certification: band 7), **French**: limited working proficiency, **German**: basic, **Czech & Slovak**: native.

Programming

Proficient in **Python** and **SQL**. Comfortable with **R**, **PySpark**, SAS, **DevOps** methodology and able to work in **Agile** environment. Familiar with PL/SQL, T-SQL and U-SQL.

Cloud services

MS Azure: Proficient with storage (blob, ADLS), Databricks, Data Factory. Comfortable with CosmosDB, SQL DB, Azure Synapse, App services, Azure DevOps.

Databases

AWS: Comfortable with EC2, S3, EMR, Redshift, CodeCommit. Familiar with IAM, ECB, SageMaker. Proficient with **Kubeflow** and **PySpark**. Familiar with k8s, docker, **ElasticSearch**, HDFS.

Proficient in **model deployment**, monitoring and serving.

Machine Learning

Supervised and unsupervised learning, dimensionality reduction, model selection, **neural networks** (Keras, Tensorflow), anomaly detection, **recommender systems**, NLP, **Computer Vision**, web scraping, **statistics**.

Presentations

Efficient programming in PySpark

Brussels, Belgium

TRAININGS

05/2021

- Prepared a training session for our clients on how to efficiently use **PySpark**.

Data Science Leuven Meetup

Brussels, Belgium

INTRODUCTION TO COBRA

09/2021

- Presented methodology and live demo of our recently open-sourced Python library **Cobra** on Data Science Leuven Meetup ([video](#))

LinkedIn

Brussels, Belgium

INTRODUCTION TO RECOMMENDER SYSTEMS

04/2020

- Written series of three articles on LinkedIn about methodology of recommendation systems [part one](#), [part two](#), [part three](#)

Hobbies

WHAT DO I DO IN MY FREE TIME?

- I love to read. A lot. Mostly classical novels, science fiction and books about science. Take a look what I have read in the [past few years](#).
- Second addiction is surely strength training and calisthenics. I can do all kinds of **fun stuff**.
- And tea, which is basically my fuel. Huge fan of Japanese sencha and oolongs from Taiwan.