

Borys

Senior Data Science Engineer

SUMMARY

Certified Data Scientist with a strong focus on NLP, CV, and Recommender Systems backed by 4 years of commercial experience. Proficient in Python with a rich toolset including Pandas, numpy, TensorFlow, and Keras. Possesses a solid track record in building products from scratch and devising innovative solutions with machine learning and data processing methodologies. Hands-on experience in deploying scalable solutions using Kubeflow, Docker, and CI/CD practices, complemented by proficiency with various databases such as MySQL and BigQuery. With a Bachelor's and Master's degrees in Cybersecurity Engineering, and continued education via a PhD, the engineer exemplifies a deep understanding of computer science fundamentals and data science trends. This technical expertise, combined with domain knowledge in e-commerce and network security, distinguishes the potential candidate as a valuable asset for fostering growth and innovation in technology-driven environments.

TECHNICAL SKILLS

| | |
|---|--|
| Main Technical Skills | UML, GCE, MVC, AWS ML (Amazon Machine learning services), Python |
| Programming Languages | Python |
| AI & Machine Learning | AWS ML (Amazon Machine learning services), Faiss, GPT, GPT-4o, Kubeflow, LangChain, OpenAI, OpenCV, Spacy, TensorFlow, Xgboost |
| Python Frameworks | FastAPI, Flask |
| Python Libraries and Tools | Matplotlib, NLTK, Plotly, PySpark, SciPy, TensorFlow |
| Data Analysis and Visualization Technologies | Kibana, Prophet, Tableau |
| Databases & Management Systems / ORM | AWS Redshift, Bigtable, ELK stack (Elasticsearch, Logstash, Kibana), MySQL, SQLAlchemy |
| Amazon Web Services | AWS ML (Amazon Machine learning services), AWS Redshift |
| Google Cloud Platform | GCE, GCP AI, Google BigQuery |
| Message/Queue/Task Brokers | Apache ActiveMQ |
| Deployment, CI/CD & Administration | CI/CD, Jenkins |

| | |
|---|---|
| Virtualization, Containers and Orchestration | Docker |
| SDK / API and Integrations | FastAPI, Google API |
| Version Control | Git |
| Logging and Monitoring | Logstash |
| Methodologies, Paradigms and Patterns | MVC, UML |
| Other Technical Skills | Chat GPT, ChurnZero, Python Requests Library, WebSphere App Serve |

WORK EXPERIENCE

Data Scientist (CV), E-commerce Data Analysis and Visualization

Duration: July 2021 - present

Summary: Conducted in-depth data analysis for E-commerce services, created interactive dashboards with Tableau, and implemented a custom ETL solution for data infrastructure enhancement.

Responsibilities: Comprehensive data analysis, ETL solution development, interactive dashboards creation

Technologies: Python, ChurnZero, MySQL, AWS RedShift, Requests, SQLAlchemy, Tableau

Data Scientist (CV), NLP Chat-bot and Product Recommendation

Duration: July 2021 - present

Summary:

- Implemented a Chat-bot for user interaction and product recommendation utilizing OpenAI GPT models including GPT-3 and GPT-4
- engineered a vector search and an API within a Docker container

Responsibilities: Preprocessing dataset, prompt engineering, vector search implementation, API creation

Technologies: Python, OpenAI, GPT-3, GPT-4, Chat GPT, Langchain, Docker

Data Scientist (CV), OpenAI Model Fine-Tuning for Text Classification

Duration: July 2021 - present

Summary: Implemented OpenAI model fine-tuning for text classification, developed an API for model updating and serving.

Responsibilities: Training data creation, dataset preparation, model fine-tuning, API development

Technologies: Python, OpenAI, GPT-3 (davinci, curie, babbage, ada)

Data Scientist (CV), OpenAI Chat-like Interaction Research

Duration: July 2021 - present

Summary: Researched and created proofs of concept for OpenAI models' chat-like discussion capabilities, used few-shot learning and fine-tuning approaches.

Responsibilities: Training data engineering, model fine-tuning, inference testing

Technologies: Python, OpenAI, GPT-3, GPT-4, Chat GPT, Langchain



Data Scientist (CV), Logistics Route Optimization

Duration: July 2021 - present

Summary: Conducted exploratory data analysis and deployed machine learning solutions for optimizing logistic routes via statistical and ML models.

Responsibilities: Data engineering, EDA execution, groups division for target data, data visualization research

Technologies: Python, Pandas, SciPy, Sklearn, Matplotlib, Plotly, Prophet, BigQuery

Data Scientist (CV), Real Estate Image Classification

Duration: July 2021 - present

Summary: Resolved multi-label classification tasks for room type and features identification from images in the real estate industry.

Technologies: Python, Tensorflow, Google Vision AutoML, Image multi-label classification, Docker, Flask

Data Scientist (NLP), Semantic Text Similarity Service

Duration: July 2021 - present

Summary: Developed a Semantic Text Similarity service to assist with the reduction of localization costs by grouping semantically similar strings prior to translation.

Technologies: Python, Tensorflow, Docker, USE, FastAPI, SQL, tf-serving

Data Scientist (NLP), Morphology Analysis Service

Duration: July 2021 - present

Summary: Developed a morphology analysis service covering POS-tagging, NER, lemmatization, and glossary extraction to maintain consistency in translations.

Technologies: Python, Spacy, nltk, Docker, Flask

Data Scientist (NLP), Translation Alignment Service

Duration: July 2021 - present

Summary: Created a service for aligning translations without identifiers to the source strings in localization management platforms.

Technologies: Python, Tensorflow, USE, Clustering, Flask, Docker, tf-serving

Data Scientist (Recommender systems), Online Retail Product Recommendations

Duration: July 2021 - present

Summary: Developed product recommendation systems for online retail using implicit feedback, performed A/B testing, and developed a model serving application.

Responsibilities: Recommendation system development, model offline and online evaluation, model serving app development

Technologies: Python, GCP, Tensorflow, Kubeflow

Data Scientist, SIEM System Enhancement

Duration: January 2020 - July 2021

Summary: Enhanced the Company's SIEM system with scalable and supportable software, including the development of ML models for network attack detection and feature engineering.

Responsibilities: Unsupervised and supervised ML models development, feature engineering, model training and evaluation

Technologies: Python, ELK stack, scikit-learn, XGBoost



EDUCATION

- **Bachelor's Degree in Cybersecurity Engineering**
Ternopil Ivan Puluj National Technical University
- **Master's Degree in Cybersecurity Engineering**
Ternopil Ivan Puluj National Technical University
- **Ongoing PhD in Computer Science**
Ternopil Ivan Puluj National Technical University

CERTIFICATION

- **Kyivstar Big Data School 4.0**
Program of the school introduces students to major concepts and techniques of data science process: predictive analytics and machine learning at scale, big data tools and technologies, basics of business analytics
2019
- **Data Science Camp Offline ML course at SmartInsight**
2021
- **Introduction to Recommender Systems: Non-Personalized and Content-Based**
Coursera
2022
- **Nearest Neighbor Collaborative Filtering**
Coursera
2022
- **Convolutional Neural Networks in TensorFlow**
Coursera
2022
- **Natural Language Processing with Classification and Vector Spaces**
Coursera
2023
- **Natural Language Processing with Probabilistic Models**
Coursera
2023

