

Viktoria T.

Data Science Engineer

SUMMARY

Data Science engineer with over 3 years of practical commercial experience in Natural Language Processing (NLP), Computer Vision (CV), and Recommender Systems. Available skills in data analysis using machine learning approaches to satisfy business needs, problem-solving, and other tasks in this sphere. A person, focused on obtaining the best results, using all knowledge and skills. Friendly and ready to help the team complete tasks and solve certain problems.

SKILLS

Main Technical Skills	Python, Computer Vision (CV), Pandas, ML, AI
AI & Machine Learning	Deep Learning, Keras, Kubeflow, Mlflow, NumPy, OpenCV, PyTorch, Scikit-learn, Spacy, TensorFlow
Python Frameworks and Libraries	Keras, Matplotlib, NLTK, NumPy, Plotly, poetry, PyTorch, Scikit-learn, SciPy, Streamlit, TensorFlow
Databases & Management Systems / ORM	Google BigQuery, MySQL
Amazon Web Services	AWS RT
Google Cloud Platform	GCP Storage, Google BigQuery
Deployment, CI/CD & Administration	CI/CD, Jenkins
Virtualization, Containers and Orchestration	Docker
Version Control	Git
SDK / API and Integrations	Payment Gateways
Scripting and Command Line Interfaces	Regex
Third Party Tools / IDEs / SDK / Services	Sublime Text

Other Technical Skills	argparse, Custom API, Deep Learning (DL), DVC, Hugging Face, Kubeflow for ML pipelines, Label Studio, MMCV, NLP, ONNX, Recommender Systems, tf-serving, Voxel51, YOLO
-------------------------------	---

WORK EXPERIENCE

Data Scientist (CV), Person Detection and Face Recognition

Duration: Duration: 1 year 1 months

Summary: The main goal was to classify people focusing on their actions and to recognize specific persons using their faces.

Responsibilities:

- Creating and engineering training data for a model.
- Implemented dataset preparation and model fine-tuning code for model evaluation.
- Evaluating different approaches to data preparation.
- Developing pipeline steps for the training model.
- Implementing new pre-trained models

Technologies: Python, Pandas, Pytorch, YOLO, RT-DETR

Data Scientist (NLP), PDF Text and Table Extraction

Summary: Extracting textual information from non-editable PDFs for quick collection and analysis. The task was related to recognizing and classifying text and tables in the picture using the "AWS Textract" service.

Technologies: Python, Matplotlib, Plotly, NumPy, Pandas, OpenCV, Regex, Spacy, AWS Textract

Duration: 4 months

Data Scientist, Online Retail Product Recommendations

Summary: Developed product recommendations system for online retail with analysis of historical user data. Developed a pipeline to generate popular items based on time, price, etc.

Responsibilities:

- Developed a recommendation system using implicit data
- Evaluated model offline and online (A/B tests) scores
- Developed model serving app and evaluated its performance

Technologies: Python, GCP, Tensorflow, Kubeflow

Duration: 6 months

Data Scientist (CV), Infrastructure Log Analysis

Summary: System for automated analysis of infrastructure logs which are row text data to discover the associated groups of resources that generate the logs based on extracted tags, ids, names, and other recognized entities.



Responsibilities:

- Using SQL-like databases for data extraction
- Extracting data from SQL-like databases based on specific task queries.
- Filtered and cleaned data to ensure accuracy and relevance.
- Merged multiple datasets into a single dataset to optimize extraction time and resources.
- Creating and filtering data using different patterns.
- Applying K-modes model for clustering words
- Preparing results taking into account the requirements of the customer

Technologies: Python, Matplotlib, Plotly, NumPy, Pandas, OpenCV, Regex, Spacy, AWS Textract

Duration: 8 months

Data Scientist (CV), Game Reward Management System

Summary: System for managing custom winnings during a game using user information such as user level, VIP rank, user wallet, etc. and generated probabilities for reward items.

Responsibilities:

- Implemented economy manager for receiving reward items.
- Created a simulation of the game using mock user data.
- Created pytests for verifying created algorithms.

Technologies: Python, Pandas, NumPy, SciPy, dataclasses

Duration: 4 months

EDUCATION

- **Bachelor's degree in System analysis**
2019 - 2023

CERTIFICATION

- **Data Science Camp Offline ML course at SmartInsight**
2021
- **Introduction to Data Science in Python**
Coursera
2020
- **Applied Machine Learning in Python**
Coursera
2020
- **Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning**
Coursera
2022



- **Convolutional Neural Networks in TensorFlow**

Coursera

2022

- **Data Science Methodology**

Coursera

2022

- **Google Cloud Big Data and Machine Learning Fundamentals**

Coursera

2023

