Taras S.
Middle Data Science Engineer

SUMMARY

Certified Data Scientist bringing an 3+ year commercial expertise, particularly in NLP, CV, and 3D modeling, underpinned by a solid foundation in machine learning and data processing. Commands an advanced proficiency in Python, complemented by adeptness in JavaScript and NodeJS. Demonstrates in-depth experience with Python libraries such as Pandas, numpy, and PyTorch, and excels in deploying solutions with Git, Docker and API development. Proven track record in industries such as SEO and logistics, leveraging statistical models and ML solutions to drive data insights and optimization. Renowned for transforming complex requirements into scalable algorithms and models, with academic credentials in Mechanical and Computer Engineering.

TECHNICAL SKILLS

Main Technical Skills	Python (3 yr.), Node.js, Pandas, NumPy, ML (2 yr.)
Programming Languages	JavaScript, Python (3 yr.)
AI & Machine Learning	AI, AWS ML (Amazon Machine learning services), Computer Vision, Deep Learning, GPT, Keras, NLP, NumPy, OpenCV, PyTorch, Scikit-learn, Spacy, Tensorboard, TensorFlow
Python Frameworks	Django, Flask
Python Libraries and Tools	Keras, Matplotlib, NLTK, NumPy, Pandas, Plotly, PyTorch, Scikit-learn, SciPy, TensorFlow, Trimesh
JavaScript Frameworks	Node.js
Data Analysis and Visualization Technologies	Logistic regression, ML (2 yr.), Pandas, Prophet
Databases & Management Systems / ORM	MongoDB, MySQL, SQL
Amazon Web Services	AWS ML (Amazon Machine learning services), AWS S3
Google Cloud Platform	Google BigQuery
UI/UX/Wireframing	3D Modelling
Virtualization, Containers and Orchestration	Docker, Kubernetes
Version Control	Git
Other Technical Skills	Computer Vision (CV), Custom API, Deep Learning (DL), Open3d, Unreal Engine (1 yr.)

WORK EXPERIENCE

Data Scientist (Computer Vision)

Duration: 1 year 6 months
Summary:

- Researched and implemented pipeline for synthetic image transformation using noise generation by simulating noise characteristics from cameras. Implemented two approaches parametric and non-parametric based on statistical analysis.
- Development of a system for reprojection of 3D dog models based on images to display animal animations on a device. Utilizing an architecture based on PyTorch and a custom-collected dataset. The process involves obtaining keypoints and segmentation from heatmaps, processing statistical components, calculating rotations, and generating the 3D dog model.

Responsibilities:

- Rewrite the code from PyTorch implementation to Keras
- Conduct separate parts of whole architecture of the Barc
- Ensure compatibility of a new model with custom dataset
- Train pretrained models based on in-house dataset
- Validation of in-house dataset
- Train and evaluation of all pipeline model
- Visualisation of both final and intermediate results of the pipeline model

Data Scientist, Logistics Route Optimization

Duration: 6 months
Summary: Completed data engineering to improve quality of data (such as creating for filling missing data, etc). Exploratory data analysis and machine learning solutions aimed to determine the most optimal route considering port characteristics. It includes statistical models, regression models, time series analysis, recommendations, and various custom approaches.

Responsibilities:

- Used SQL-like databases for data engineering as well as data extraction
- Executed the EDA of port-related data
- Analyzed data for the insights and patterns, as well as presented results of investigation
- Researched the approaches to visualize the data insights for customers
- Implemented ML solutions for estimating the port cost
- Used SQL-like databases for data engineering as well as data extraction
- Executed the EDA of port-related data
- Analyzed data for the insights and patterns, as well as presented results of investigation
- Researched the approaches to visualize the data insights for customers
- Implemented ML solutions for estimating the port cost

Technologies: Python, Pandas, SciPy, Scikit-learn, Matplotlib, Plotly, Prophet, BigQuery

Data Scientist, Search Engine Optimization (SEO)

Duration: 1 year
Summary: Development of custom solutions for clustering keywords by semantic similarity and other characteristics, including web application for a demo based on Django.
Technologies: Python, Pandas, Numpy, Scikit-learn, Matplotlib, spaCy, USE, GPT3, Django, Docker, Kubernetes, AWS S3

EDUCATION

- **Bachelor's Degree in Mechanical Engineering**
- **Master's Degree in Computer Engineering, ongoing**

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
- **Sequences, Time Series and Prediction**
  Coursera course.
  2023
- **Convolutional Neural Networks in TensorFlow**
  Coursera course.
  2023
- **Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning**
  Coursera course.
  2023
- **Natural Language Processing in TensorFlow**
  Coursera course.
  2023