Iosif Guzeev

Machine learning engineer with more than two years of experience

: IosifGuzeev@outlook.com
: +79521760451
: linkedin.com/in/iosif-guzeev-68290b1ab/

WORK EXPERIENCE

Middle Machine Learning engineer Aug 2022 - Present

Okko

- Made CatBoost ranker model for search engine. Including model engineering, features calculation AirFlow pipeline and model inference with calculations optimization due to latency limits. Increasing average watch time by 3% and CTR by 2.5%.
  Stack: CatBoost, Optuna, Shap, AirFlow, Redis, Flask, dynaconfig.
- Proposed and led a search personalization project by taking into account user interactions with items in the past. Increasing UX and search time before first click.
- Made a personal recommendation ML model for "popular"section based on user and item features. Improving UX by providing more relevant items in search section.
  Stack: Catboost, pandas.
- Proposed and led a project to cache requests to the model. The 90th quantile of service response time were reduced by 15%, which allowed us to use more complex models in production.
- Made a model for picking a similar titles based on it's textual info and LFM scores. Developed ETL pipeline and inference for given model. As result model is used for generating candidates in film's "similar"section, increasing amount of clicks in this section significantly.
  Stack: LangChain, chromaDB, LightFM, AirFlow.
- Made a research of methods for predicting product metrics changes based on the offline metrics values. Produced model that found positive connections between ranking metrics and our product metrics based on econometric analysis methods. Correlated validation metrics with product metrics were found. Produced methodology of it's correlation evaluation.
  Stack: Trino, Clickhouse, Scipy.

Junior Data Scientist Sep 2021 - Aug 2022

PerfectArt

- Built ETL + ML pipeline to find best bucket embedding algorithm in e-commerce project. TF-IDF, spacy, starspace and TF-IDF + Neural network algorithms were applied to concatenated textual description in the customers bucket.
  Stack: DVC, Sklearn, PyTorch, boto3.
- Made fuel consumption research for private tracking company. Raw data with information about fuel consumption per driver and region were analyzed. Data was given as PostgreSQL database. As result, automatically executed Jupyter Notebook with plots and tables were made.
  Stack: PostgreSQL, Psycopg2, Pandas, Jupyter Notebook, Matplotlib.
- Made ETL pipeline for processing raw signal measurements from custom medical device. This pipeline includes averaging and FFT.
  Stack: DVC, Pandas, Scipy, Matplotlib, Seaborn.
- Finished successfully standalone software project without any violation of deadlines. Custom firmware for NXP controller and Labview interface were made.
  Stack: IAR Embedded Workbench, LabView.

SKILLS

- **Languages**: Russian (Native), English (upper-intermediate)
- **Software Languages & Tools**: Python, SQL (Trino, Clickhouse), Git, C/C++, Docker, DVC, Airflow; Experienced in Matlab, LabView, ChromaDB, Amazon services (S3, Lambda)
- **Data Analysis**: Pandas, Scipy, Matplotlib, Seaborn, Plotly, Splunk
- **Machine learning**: Sklearn, CatBoost, PyTorch, Langchain, LightFM
- **Mathematical background**: Probabilities, Statistics, Numerical methods, Analysis, Discrete mathematics, A\B Testing
- **Project management**: Git, Bitbucket, Jira, Tempo, Confluence, Draw.io

EDUCATION

Master's degree in Big data analysis 2022 - 2024
Tomsk state university | Tomsk, Russia

- **Courses**: artificial intelligence, virtualization, augmented reality, post-relational databases, digital signal processing.
- **Thesis**: Feature selection in Raman spectroscopy for cell organelles classification (in progress).

Bachelor's degree in Applied mathematics and computer science 2017 - 2021
Tomsk state university | Tomsk, Russia

- **Courses**: analysis, probabilities, statistics, numerical methods, architecture design, discrete mathematics, differential equations, computer architecture.
- **Conferences**: Tomsk State University International conference (best participant in ML section, article was published)
- **Thesis**: Group method of data handling algorithms. Analysis and comparison. (can be downloaded from university site)