Pavel Repnikov

Location: Moscow, Russia

GitHub | Email: prepnik@mail.ru | Mobile: +79888639559

Technical Skills

Languages : C++, Python, SQL

Frameworks : Unreal Engine, Unity, Pytorch

Libraries : numpy, pandas, PySpark, Greenplum, XGBoost, LightGBM, CatBoost, numba, sklearn, geopandas, SciPy, PySR, PyClustering,Optuna

Work Experience

Data Scientist Intern Oct 2023 – Present

Sberbank, Risk Modeling Department *Moscow, Russia*

- Geolocation data. Geolocation data in fraud prevention
- Behavior modeling. Probabilities of events as a new way of explaining incidents
- Working with classic tabular data
- Building datasets from multiple data sources
- Time series forecasting

Education

Lomonosov Moscow State University 2022-2024

MSc in Physics,Chair of Mathematical Modeling and Computer Science *Moscow, Russia*

Lomonosov Moscow State University 2018-2022

BSc in Physics,Chair of Mathematical Modeling and Computer Science *Moscow, Russia*

Completed projects

Forecasting global population dynamics *Python, numba, geopandas* code

- Partial differential equations as a way to predict the population on the globe
- Modification of the classical formulation of the problem taking into account spatial components
- A solution on a set of arbitrary shape

Credit Scoring on a synthetic dataset *Python, XGBoost, LightGBM, CatBoost, PyClustering, SciPy* code

- The divide and conquer principle. Building models independently for different years
- Automatic feature generation
- Testing statistical hypotheses
- Clustering of tabular data

Furniture object detection *SQL,C#, Unity, Python, Pytorch* code

- Creating a synthetic dataset using Unity
- Object detection finetuning using Pytorch

Bayesian Decision Making as a Theoretical Basis for a New Look at Fuzzy Logic Control *Python, Pytorch* code

- Creating a new machine learning white-box model from scratch
- Creating a fuzzy inference system based on statistical inference
- Solving a system of integral equations using Pytorch

Machine learning of noise filtering of vibroacoustic linearly distributed sensor data *Python, TensorFlow* code

- Creating an optimal signal filter for recognizing different types of activity

Galton board modeling *C++,Qt* code

- Demonstration of the central limit theorem
- Creating a desktop application from scratch
- Embedding an engine for physical simulation

Projects in development

Adaptive metabolic model Aug 2023 – Present

- Monte Carlo simulation
- Time series clustering

Adaptive control system with fuzzy logic based on Bayesian inference Aug 2022 – Present

- Cross-entropy method for reinforcement learning
- Creating a greedy optimization algorithm for physical simulation
- Creating an analogue of the gradient descent algorithm in the function space