

# MAKSIM KIRYAKIN

+7 (916) 344-6061 ◇ Russia, Moscow

[maksim.kiryakin@gmail.com](mailto:maksim.kiryakin@gmail.com) ◇ [github.com/MaximKiryakin](https://github.com/MaximKiryakin) ◇ [linkedin.com/in/your-profile](https://linkedin.com/in/your-profile)

## SUMMARY

**Data Scientist with 1.5 years of experience** in classical machine learning, econometrics, and mathematical programming. Specializing in scalable ML solutions and financial analytics. Delivered clustering pipeline for 120M+ users and deployed NPV models to production, driving data-driven decision-making. Proficient in Python (PySpark, Pandas), SQL, and end-to-end ML pipelines.

## EDUCATION

**Lomonosov Moscow State University (Master's degree)** September 2024 - Present  
Faculty of Computational Mathematics and Cybernetics (*GPA: 4.4/5*).

**Lomonosov Moscow State University (Bachelor's degree)** September 2020 - June 2024  
Faculty of Computational Mathematics and Cybernetics (*GPA: 4.3/5*).

## WORK EXPERIENCE

**Sber** May 2024 - Present  
Data Scientist, Investment Expertise Directorate (Finance Division)

- Architected and productionized a scalable customer clustering pipeline for 120M+ profiles using PySpark.
- Led R&D to define, implement a new bank-wide metric, including statistical evaluation and experiments.
- Developed, validated, and deployed ML and econometric model to estimate product NPV; integrated them with data pipelines and established monitoring, maintenance, and business-acceptance procedures to ensure robustness and interpretability.
- Stack: Python (PySpark, Pandas, Scikit-learn), SQL, Hive, Git, Excel.

**Sber** August 2023 - March 2024  
Intern, Balance Sheet Structure Department (Finance Division)

- Refactored and modularized legacy Python prototypes, added unit tests and documentation to improve maintainability and accelerate onboarding.
- Integrated a pre-trained ML model into the production pipeline, improving target prediction quality by 30% on validation data and ensuring reproducible inference.
- Optimized the main computation (profiling-driven algorithmic refactor and vectorization), achieving a 3× speedup and reduced memory footprint.
- Developed automated ETL workflows to extract, validate and preprocess large datasets from HDFS using PySpark/Python, with logging and error handling for reliable runs.
- Stack: Python (PySpark, Pandas, Scikit-learn), SQL, Hive, Excel.

## SKILLS

**Programming:** Python, SQL, C/C++  
**Data & ML:** PySpark, Pandas, NumPy, SciPy, Scikit-learn, PyTorch, SQL, Hive  
**Tools:** Git, Excel  
**Concepts:** Feature engineering, clustering, time-series, econometrics, model deployment

## PROFESSIONAL DEVELOPMENT

- *Apache Spark for Data Analysis Tasks* — New Professions Lab [Certificate \(with honors\)](#)
- *Methods of Data Analysis and Machine Learning* — Sber University [Certificate \(with honors\)](#)
- *Mathematical Models in Investment Banks* — Deutsche Bank Technology Center