Ekaterina Bremel

New York, NY | (646) 945-9649 | bremelket@gmail.com | LinkedIn | GitHub | Portfolio

Summary

Computer Science student (**4.0 GPA, Dean's List**) with hands-on ML experience delivering measurable business impact: optimized data workflows on large-scale datasets **achieving 35% efficiency gains** in data workflows, developed predictive models with **23% error reduction**, and enhanced LLM performance to **85% accuracy through prompt engineering**. Skilled in data pipelines, **A/B testing**, and collaborating cross-functionally to translate analytical insights into product decisions.

Technical Skills

Languages: Python, SQL, C++

ML Frameworks & Libraries: Scikit-learn, TensorFlow, Keras, LightGBM, CatBoost, Pandas, NumPy

ML Techniques: Regression, Classification, Clustering, Ensemble Methods (Random Forest, CatBoost, LightGBM), Hyperparameter Tuning, Cross-Validation

Deep Learning & NLP: CNNs (ResNet50), TF-IDF, Word Embeddings, NLTK, Text Classification, Computer Vision

Data Analysis & Visualization: A/B Testing, EDA, Feature Engineering, Statistical Analysis, Matplotlib, Seaborn, Data Storytelling

Tools & Additional: GitHub, Jira, Jupyter Notebook, Large Language Models (LLMs), Prompt Engineering, Cross-team Collaboration, Stakeholder Communication

Education

St. Francis College — Expected Graduation: December 2026

Bachelor of Science in Computer Science — Brooklyn, NY

- **Dean's List: All Semesters, Varsity Tennis Team - GPA: 4.0/4.0**
- Courses: Machine Learning, Software Engineering, Data Structures & Algorithms, Database Management Systems

Yandex School of Data Analysis

Data Science Certificate - GPA: 4.0/4.0

- Completed comprehensive ML curriculum including regression, classification, NLP, and computer vision

Work Experience

HumanKind OS — Apr 2025 - Present

Machine Learning Engineer (Part-time) — New York, NY

- Analyzed wearable device data to identify user engagement patterns, translating insights into product recommendations that informed feature prioritization for beta launch
- **Built data pipelines and dashboards** tracking key metrics, enabling product and engineering teams to monitor early user behavior and iterate on core features
- **Designed A/B testing framework** for validating feature hypotheses, analyzing user behavior patterns to guide product decisions during beta development phase
- **Developed LLM-based syste**m to predict user behavior patterns and generate personalized wellness recommendations for beta launch
- **Collaborated across all teams** (founders, engineering, design, marketing) to integrate data insights into product strategy, shaping MVP roadmap and go-to-market decisions

Sberbank- Eastern Europe's largest bank, $500B+ assets, serves 108M+ clients — Feb 2024 - Apr 2025

Data Scientist

- Optimized data processing workflows by **35% through Python-based pipeline optimization** (Pandas, NumPy), eliminating redundancy and enabling real-time analytics for operations and finance stakeholders
- **Improved predictive model performance by 23% error reduction**, enabling operations managers to make better decisions that increased service delivery efficiency and customer satisfaction
- Increased LLM-based customer feedback classifier to **85% accuracy via prompt engineering and post-processing**, improving automated classification used by product and marketing teams
- Analyzed large-scale customer data using Python to identify behavioral patterns and usage trends, delivering visualizations that shaped marketing segmentation strategy and informed product recommendations
- Collaborated across product, marketing, and engineering teams using Jira and Slack to translate business requirements into data solutions and present technical findings to non-technical stakeholders

Projects

Steel Temperature Prediction for Energy Optimization | Source Code — *Python | Scikit-learn| Machine Learning | CatBoost*

- Developed regression model forecasting steel temperatures to optimize energy usage, achieving MAE of 4.05°C and enabling 15-20% reduction in electricity consumption for manufacturing operations

Toxic Comment Classification for Content Moderation | Source Code — *Python | NLTK |Matplotlib | Logistic Regression*

- Built NLP classifier for automated content moderation using TF-IDF, achieving F1 score of 0.78 (above 0.75 threshold), reducing manual review workload by 60% while maintaining healthy online communities

Customer Age Estimation using Computer Vision | Source Code — *TensorFlow | Computer Vision | ResNet50 | CNN*

- Created CNN using transfer learning to estimate customer age from photos, achieving MAE of 5.89 years, enabling targeted marketing campaigns and supporting age-restricted product compliance without manual verification

Gold Recovery Prediction for Ore Processing Optimization | Source Code — *Pandas | Random Forest | NumPy | Cross-Validation*

- Built predictive model for mining efficiency optimization, achieving sMAPE of 8.97%, enabling mining operations to avoid processing unprofitable batches and optimize resource allocation for improved margins