

# Ruslan Borisenco

---

email: ruslanborysenko@gmail.com

tel: 730 837 748

linkedin: <https://www.linkedin.com/in/ruslan-borisenco-60a56319b/>

github: <https://github.com/ruslanborysenko-lab>

Junior Data Scientist specializing in end-to-end ML pipeline development, computer vision, NLP, and cloud-based solutions. Experienced in building production-ready AI systems from data exploration through deployment, with focus on MLOps, LLM integration, and business intelligence dashboards.

Develop and deploy machine learning models for diverse applications including OCR invoice processing, conversational AI chatbots, clustering algorithms for recommendation systems, and predictive analytics. Design interactive data visualization dashboards for survey analytics and business metrics. Implement semantic search systems with vector embeddings and optimize warehouse operations using data-driven approaches.

Proficient in Python ecosystem (pandas, scikit-learn), cloud platforms. Strong foundation in exploratory data analysis, statistical modeling, and SQL database management. Demonstrated ability to translate business requirements into technical solutions with measurable impact.

## Projects

### Half Marathon Time Prediction - ML & NLP

ML-powered web app predicting half marathon finish times from 5km results using PyCaret regression models ( $R^2>0.85$ ) trained on 8,000+ runner dataset. Features GPT-4o-mini NLP for natural language input parsing, cloud model storage on Digital Ocean Spaces, Langfuse LLM observability, and comparative analytics with 14 age/gender categories.

- **Skills:** PyCaret, OpenAI GPT-4o-mini, Langfuse, Digital Ocean Spaces, boto3, Streamlit, pandas, matplotlib
- **Results:**  $R^2>0.85$  model on 8,000+ samples, cloud deployment with LLM monitoring, natural language NLP parsing

### TransStream - Warehouse Workforce Optimization

Enterprise-grade workforce optimization system combining GPT-4o AI with business intelligence analytics. Analyzes 1,500m<sup>2</sup> warehouse operations (May-Sept), optimizes 5 employee roles using formula-based calculations, forecasts Oct-Dec staffing with linear regression. Features 10+ interactive dashboards, executive KPIs, and real-time optimization. Largest project (1,369 lines of code).

- **Skills:** OpenAI GPT-4o, Prompt Engineering, Linear Regression, Streamlit, Plotly, pandas, Business Intelligence
- **Results:** 60% Loader reduction (10→4), executive dashboard with 4 KPIs, 3-month forecast, 10+ visualizations

## ML Friend Finder - Clustering Application

Machine learning application using K-Means clustering to match users with similar interests. Built with PyCaret for automated ML pipeline and Streamlit for web interface. Trained on 229 survey respondents, the model segments users into 8 groups. Features real-time prediction, interactive visualizations, and Polish localization.

- **Skills:** K-Means Clustering, PyCaret, Unsupervised Learning, Streamlit, Plotly, Feature Engineering
- **Results:** 8 user segments, 5→21 feature transformation, Silhouette Score 0.195, <1s prediction with caching

## AI Voice Notes - Semantic Search Application

Advanced voice notes app with AI-powered transcription and semantic search. Built with OpenAI Whisper for speech-to-text, text-embedding-3-large for 3072-dim vectors, and Qdrant for similarity search. Features browser audio recording, MD5 caching, and cloud deployment. Find notes by meaning, not just keywords.

- **Skills:** OpenAI Whisper, Vector Embeddings, Qdrant, Semantic Search, Streamlit, Audio Processing
- **Results:** 3072-dim vectors, <1s semantic search, MD5 caching for cost optimization, cloud deployment

## Interactive Survey Data Analysis Dashboard

Comprehensive data visualization web app analyzing 140 student profiles across 11 variables. Built with Streamlit and Plotly, featuring interactive charts, correlation analysis, and automated insights. Implemented smart category ordering, missing data handling, and multi-stage analytics pipeline for educational program optimization.

- **Skills:** Streamlit, Plotly, pandas, Statistical Analysis, Data Visualization, Correlation Analysis
- **Results:** 11 analytical modules, 3-stage pipeline, 1,030 lines of code, regex age interval parser, comparative analytics

## AI Chatbot Web Application

Production-ready web chatbot built with Streamlit and OpenAI GPT-4o/GPT-5. Features multi-conversation management, customizable AI personalities, persistent JSON-based storage, and token usage tracking. Implemented session state management and 20-message memory window for cost optimization.

- **Skills:** Streamlit, OpenAI GPT-4o, Session State Management, JSON, API Integration
- **Results:** Unlimited conversations, 20-message memory window (~60% token reduction), token usage tracking

## AI-Powered Invoice Data Extraction System

Automated data processing pipeline using OpenAI GPT-4o Vision API to extract structured information from invoice images. Integrated multi-source data (SQLite, CSV, images) into unified dataset. Achieved 100% automation of manual invoice processing using LLM with Pydantic validation.

- **Skills:** OpenAI GPT-4o Vision, Pydantic, pandas, SQLite, ETL, Data Integration

- **Results:** 100% automation, processed 10 invoices with 16+ line items, integrated 3 data sources, type-safe validation

## Titanic Disaster Survival Analysis

Comprehensive exploratory data analysis of 1,310 Titanic passengers examining survival patterns by class, gender, age, and family size. Analyzed correlations between socioeconomic status and survival rates, identified rescue priorities (73% female survival vs 20% male), and explored lifeboat allocation patterns across passenger classes.

- **Skills:** pandas, matplotlib, Statistical Analysis, Missing Data Analysis, Correlation Analysis
- **Results:** Identified 73% vs 20% survival disparity, handled 77% missing cabin data, analyzed 27 lifeboats

## Iris Species Classification - EDA

Comprehensive exploratory data analysis of the classic Iris flower dataset covering three species (setosa, versicolor, virginica). Performed statistical analysis on 150 samples with 4 morphological features, including correlation analysis, distribution studies, and species differentiation using pandas visualization.

- **Skills:** pandas, NumPy, matplotlib, Descriptive Statistics, Data Visualization
- **Results:** Identified petal length as strongest discriminator, 0.75 correlation for setosa, zero missing values

## Technical Skills

- **Programming Languages:** Python, SQL
- **Libraries and Tools:** pandas, scikit-learn, PyCaret, OpenAI API, Streamlit, Plotly, matplotlib, seaborn, Git, Jupyter Notebook, Docker
- **Databases:** SQLite, Qdrant
- **Other:** Google Cloud, Digital Ocean, Langfuse, Prompt Engineering, Vector Databases, REST APIs

## Portfolio

[Visit my portfolio here](#)

## Education

- **From Zero to AI** - Gotoit - 2025
- **Bachelor of Engineering - BE, Engineer-constructor** – Dnipropetrovsk State University, Ukraine, Faculty of physics and technology – 1993-1998

## Languages

- **English:** B2
- **Polish:** B2
- **Ukrainian:** Native