

# 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 specjalizujący się w rozwiązaniach ML end-to-end, computer vision, NLP i wdrożeniach w chmurze. Doświadczenie w budowie systemów AI gotowych do produkcji – od eksploracji danych przez wdrożenie, z naciskiem na MLOps, integrację LLM i dashboardy business intelligence.

Tworzę i wdrażam modele machine learning dla różnorodnych zastosowań: przetwarzanie faktur OCR, chatboty konwersacyjne AI, algorytmy klastrowania dla systemów rekomendacji oraz analityka predykcyjna. Projektuję interaktywne dashboardy wizualizacji danych dla analityki ankiet i metryk biznesowych. Implementuję systemy wyszukiwania semantycznego z wektorowymi embeddingami oraz optymalizuję operacje magazynowe przy użyciu podejść opartych na danych.

Biegłość w ekosystemie Python (pandas, scikit-learn), platformach chmurowych. Silne fundamenty w eksploracyjnej analizie danych, modelowaniu statystycznym i zarządzaniu bazami danych SQL. Udokumentowana umiejętność przekładania wymagań biznesowych na rozwiązania techniczne z mierzalnym wpływem.

## Projekty

### Half Marathon Time Prediction - ML & NLP

Aplikacja webowa ML przewidująca czasy półmaratonu z wyników 5km używając modeli regresji PyCaret ( $R^2>0.85$ ) wytrenowanych na 8000+ biegaczy. Integracja GPT-4o-mini NLP do parsowania języka naturalnego, przechowywanie modelu w chmurze Digital Ocean Spaces, obserwability LLM przez Langfuse oraz analityka porównawcza dla 14 kategorii wiek/płeć.

- **Umiejętności:** PyCaret, OpenAI GPT-4o-mini, Langfuse, Digital Ocean Spaces, boto3, Streamlit, pandas, matplotlib
- **Wynik:** Model  $R^2>0.85$  na 8000+ próbkach, wdrożenie w chmurze z monitoringiem LLM, parsowanie NLP języka naturalnego

### TransStream - Warehouse Workforce Optimization

System optymalizacji pracowników magazynu klasy enterprise łączący AI GPT-4o z analityką business intelligence. Analiza operacji magazynu 1500m<sup>2</sup> (maj-wrzesień), optymalizacja 5 ról pracowniczych używając obliczeń wzorowych, prognoza obsad paź-gru regresją liniową. 10+ interaktywnych dashboardów, wskaźniki KPI dla zarządu, optymalizacja real-time. Największy projekt (1369 linii kodu).

- **Umiejętności:** OpenAI GPT-4o, Prompt Engineering, Linear Regression, Streamlit, Plotly, pandas, Business Intelligence
- **Wynik:** Redukcja 60% pracowników Loader (10→4), dashboard wykonawczy z 4 KPIs, prognoza 3-miesięczna, 10+ wizualizacji

## ML Friend Finder - Clustering Application

Aplikacja machine learning używająca klastrowania K-Means do dopasowywania użytkowników o podobnych zainteresowaniach. Zbudowana z PyCaret dla zautomatyzowanego pipeline ML i Streamlit dla interfejsu webowego. Wytrenowana na 229 respondentach ankiety, model segmentuje użytkowników na 8 grup. Predykcja real-time, interaktywne wizualizacje, lokalizacja polska.

- **Umiejętności:** K-Means Clustering, PyCaret, Unsupervised Learning, Streamlit, Plotly, Feature Engineering
- **Wynik:** 8 segmentów użytkowników, transformacja 5→21 cech, Silhouette Score 0.195, predykcja <1s z cachingiem

## AI Voice Notes - Semantic Search Application

Zaawansowana aplikacja notatek głosowych z transkrypcją AI i wyszukiwaniem semantycznym. Zbudowana z OpenAI Whisper dla speech-to-text, text-embedding-3-large dla wektorów 3072-dim oraz Qdrant dla wyszukiwania podobieństwa. Nagrywanie audio w przeglądarce, caching MD5, wdrożenie w chmurze. Znajdź notatki po znaczeniu, nie tylko po słowach kluczowych.

- **Umiejętności:** OpenAI Whisper, Vector Embeddings, Qdrant, Semantic Search, Streamlit, Audio Processing
- **Wynik:** Wektory 3072-dim, wyszukiwanie semantyczne <1s, caching MD5 optymalizujący koszty, wdrożenie w chmurze

## Interactive Survey Data Analysis Dashboard

Kompleksowa aplikacja webowa wizualizacji danych analizująca 140 profili studentów w 11 zmiennych. Zbudowana z Streamlit i Plotly, zawiera interaktywne wykresy, analizę korelacji oraz zautomatyzowane wnioski. Inteligentne sortowanie kategorii, obsługa brakujących danych oraz pipeline analityczny 3-stopniowy dla optymalizacji programu edukacyjnego.

- **Umiejętności:** Streamlit, Plotly, pandas, Statistical Analysis, Data Visualization, Correlation Analysis
- **Wynik:** 11 modułów analitycznych, pipeline 3-stopniowy, 1030 linii kodu, parser interwałów wieku regex, analityka porównawcza

## AI Chatbot Web Application

Gotowa do produkcji webowa aplikacja chatbota z zarządzaniem multi-konwersacjami. Zbudowana ze Streamlit i OpenAI GPT-4o/GPT-5. Zarządzanie wieloma konwersacjami, personalizowane osobowości AI, persystencja bazowana na JSON oraz śledzenie użycia tokenów. Zarządzanie session state i okno pamięci 20 wiadomości dla optymalizacji kosztów.

- **Umiejętności:** Streamlit, OpenAI GPT-4o, Session State Management, JSON, API Integration
- **Wynik:** Nieograniczone konwersacje, okno pamięci 20 wiadomości (~60% redukcja tokenów), śledzenie użycia tokenów

## AI-Powered Invoice Data Extraction System

Zautomatyzowany pipeline przetwarzania danych używający OpenAI GPT-4o Vision API do ekstrakcji strukturalnych informacji z obrazów faktur. Integracja danych multi-źródłowych (SQLite, CSV, obrazy) w

zunifikowany dataset. Osiągnięto 100% automatyzacji ręcznego przetwarzania faktur używając LLM z walidacją Pydantic.

- **Umiejętności:** OpenAI GPT-4o Vision, Pydantic, pandas, SQLite, ETL, Data Integration
- **Wynik:** 100% automatyzacja, przetworzono 10 faktur z 16+ pozycjami, integracja 3 źródeł danych, walidacja type-safe

## Titanic Disaster Survival Analysis

Kompleksowa eksploracyjna analiza danych 1310 pasażerów Titanica badająca wzorce przeżycia według klasy, płci, wieku i wielkości rodziny. Analiza korelacji między statusem społeczno-ekonomicznym a współczynnikami przeżycia, identyfikacja priorytetów ratunkowych (73% przeżycia kobiet vs 20% mężczyzn) oraz eksploracja wzorców alokacji łodzi ratunkowych przez klasy pasażerów.

- **Umiejętności:** pandas, matplotlib, Statistical Analysis, Missing Data Analysis, Correlation Analysis
- **Wynik:** Identyfikacja dysproporcji 73% vs 20% przeżycia, obsługa 77% brakujących danych kabiny, analiza 27 łodzi ratunkowych

## Iris Species Classification - EDA

Kompleksowa eksploracyjna analiza danych klasycznego datasetu kwiatów Iris obejmująca trzy gatunki (setosa, versicolor, virginica). Przeprowadzona analiza statystyczna na 150 próbkach z 4 cechami morfologicznymi, zawierająca analizę korelacji, studia dystrybucji oraz różnicowanie gatunków używając wizualizacji pandas.

- **Umiejętności:** pandas, NumPy, matplotlib, Descriptive Statistics, Data Visualization
- **Wynik:** Identyfikacja długości płatka jako najsilniejszego dyskryminatora, korelacja 0.75 dla setosa, zero brakujących wartości

## Umiejętności techniczne

- **Języki programowania:** Python, SQL
- **Biblioteki i narzędzia:** pandas, scikit-learn, PyCaret, OpenAI API, Streamlit, Plotly, matplotlib, seaborn, Git, Jupyter Notebook, Docker
- **Bazy danych:** SQLite, Qdrant
- **Inne:** Google Cloud, Digital Ocean, Langfuse, Prompt Engineering, Vector Databases, REST APIs

## Portfolio

[Moje portfolio znajdziesz tutaj](#)

## Wykształcenie

- **Od zera do AI** - Gotoit - 2025
- **Bachelor of Engineering - BE, Engineer-constructor** – Dniepropetrovsk State University, Ukraine, Faculty of physics and technology – 1993-1998

## Języki obce

- **Angielski:** B2
- **Polski:** B2
- **Ukraiński:** Ojczysty