DARIA DUBOVSKAIA

Data Scientist
Tampa, FL | (929) 533-9032 | dubovskaiad@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

- **Data Scientist** with a track record of building end-to-end solutions in data science, machine learning, data engineering, and applied analytics for education, healthcare, and public policy
- Developed interpretable ML models (XGBoost, Random Forest, SVM) and scalable pipelines using Python and SQL, communicated insights via interactive dashboards and technical reports
- Engineered analytics systems from high-frequency sensor feeds (~1,000 readings/sec) to Python- and SQL-based dashboards, improving data reliability and operational efficiency by 20–40% across research and education environments
- Designed AI-driven feedback systems for XR simulations by integrating LLM-powered virtual tutors and real-time behavioral analytics, boosting learner engagement by 35% during pilot studies, and co-authoring an NSF-backed analytics framework for data-driven healthcare simulation research

TECHNICAL SKILLS:

Programming Languages: Python, R, SQL (PostgreSQL, MySQL)
Libraries & Frameworks: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, XGBoost, Matplotlib, Seaborn, Plotly, LangChain
Data Science & Analytics: Exploratory Data Analysis (EDA), Statistical Modeling, Hypothesis Testing, A/B Testing, Bayesian Methods, Data Visualization, SHAP Explainability, Feature Engineering, Optimization Algorithms
Machine Learning & NLP: Supervised Learning (Regression, Classification, SVM, Random Forest), Unsupervised Learning (Clustering, Dimensionality Reduction, Anomaly Detection), Predictive Modeling, Time Series Forecasting, Neural Networks (Deep Learning), Hyperparameter Tuning, Model Evaluation, LLMs, RAG, Transformer Models, Natural Language Processing, Text Mining
Data Engineering: ETL/ELT Pipeline Design, Data Cleaning & Wrangling, Data Warehousing, Apache Spark, Apache Kafka, Data Quality
Tools & Platforms: AWS (S3), Docker, Git, REST APIs, Jupyter Notebooks, Power BI, Tableau

PROFESSIONAL EXPERIENCE

Data Scientist — **Tampa, FL**
Encountive, LLC — *Feb 2025 - Present*

- Developed analytics for **30+** immersive XR training scenarios, influencing learning outcomes for **60+** nursing students in multi-institution pilot studies across **3** universities
- Optimized a room-to-VR conversion pipeline, reducing processing time by **20%** for faster deployment of training simulations
- Built Python-based data pipelines to process ~**20K** simulation interaction events, quantifying trainee performance metrics (e.g., communication clarity, effective de-escalation)
- Integrated **LLM-powered** virtual tutors and adaptive avatars to enable real-time, speech-driven learning feedback, improving learner engagement by an estimated **35%**
- Contributed to a **$150K research proposal** for the NSF Smart & Connected Health program by designing the data analytics framework to measure improvements in clinical communication training outcomes
- Reduced content QA cycle time by **30%** via automated validation tools built in **Python** and **SQL**, improving simulation reliability
- Collaborated **cross-functionally** with engineering and product teams to scale emotionally intelligent healthcare simulations for rural and under-resourced healthcare markets, ensuring model performance and data ethics compliance

Master's Research Assistant — **New York, NY**
CUNY School of Professional Studies — *Oct 2022 - May 2025*

- Completed **five end-to-end data science projects** (FEMA Aid Equity, Heart Disease Prediction, Car Insurance Risk, Airbnb Forecasting, U.S. Crime Analysis) using machine learning, data engineering, and statistical modeling on datasets totaling **900K+** records
- Developed interpretable predictive models (**XGBoost, Random Forest, SVM, Lasso**) with measurable outcomes such as identifying **77%** of underfunded households, increasing heart disease prediction recall by **18%**, and improving insurance risk model accuracy by **15%**
- Delivered results through data visualization and reporting, using Python (**Pandas, Seaborn, Plotly**), **SQL**, and **AWS** to create interactive dashboards, academic white papers, and policy recommendations

IT Support Intern — **New York, NY**
Russian School of Mathematics — *Jan 2020 - Aug 2022*

- Managed academic and technical operations for **500+ students**, streamlining scheduling, grading, and data workflows to boost overall efficiency by **25%**
- Delivered supplemental instruction in advanced mathematics (calculus, statistics, linear algebra), resulting in improved test scores and stronger quantitative reasoning
- Provided IT support for hybrid and remote classrooms, maintaining **99%** system uptime for e-learning platforms
- Developed automated reporting dashboards using **Excel** and **Python**, enabling faculty to monitor student performance metrics and resource utilization in real time

Engineering Project Lead
Bauman State Technical University — *Nov 2017 - Feb 2018*

- Designed and developed laboratory test stands for rocket engine experiments, ensuring full compliance with engineering and safety standards and enabling **20+** successful static fire tests

- Developed a Python-based high-frequency data acquisition system (~**1,000 readings/sec**) for real-time engine monitoring and anomaly detection through downstream analysis pipelines
- Optimized instrumentation through systematic calibration and statistical error analysis, reducing measurement uncertainty by **18%** and improving data reproducibility
- Collaborated with **cross-functional teams** (mechanical, electrical, data, and controls engineers) to resolve integration issues, improving overall reliability and data integrity across test campaigns
- Performed post-test data analysis and visualization using **Python** (NumPy, Matplotlib) to inform engine performance improvements and guide future design iterations

EDUCATION

CUNY School of Professional Studies, New York, NY
Master of Science in Data Science **GPA: 4.0**
New York City College of Technology, New York, NY
Bachelor of Technology in Mechanical Engineering **GPA: 4.0**
Bauman Moscow State Technical University
Master of Power Engineering in Rocket Engines Design **GPA: 4.0**