

David A. Pogrebitskiy

Boston, MA | (303) 810-6328 | pogrebitskiy.d@northeastern.edu
pogrebitskiy.com | github.com/pogrebitskiy | linkedin.com/in/davidpog

EDUCATION

Northeastern University, *Khoury College of Computer Sciences*, Boston, MA **GPA: 3.98/4.00**
Candidate for Bachelor of Science in Data Science, Mathematics Minor Expected Aug 2024
Relevant Coursework: Neural Networks, Machine Learning 2, Data Visualization, Large-Scale Storage & Retrieval, Mathematics of Machine Learning, Linear Algebra, Multi-variable Calculus, Probability & Statistics

TECHNICAL SKILLS

Languages: Python, C++, SQL, Java, HTML, JavaScript
Libraries/Frameworks: Pandas, NumPy, PyTorch, Scikit-learn, Tensorflow, Plotly, HuggingFace
Databases/Platforms: MySQL, HDFS, Airflow, Dremio, MongoDB, Redis, Neo4j, Spark, Elasticsearch

EXPERIENCE

Automated Execution Analyst Co-op Jul 2023 – Present
TD Cowen / TD Securities *New York, NY*

- Spearheaded the development of Python-based FIX Protocol message translators that were deployed on Airflow, significantly enhancing in-house analytics capabilities and meeting intricate client-specific requirements
- Developed a Plotly dashboard to enhance the interpretability of quantitative trading signals and facilitate analysis of order routing technology for enhanced decision-making
- Produced comprehensive monthly market reports utilizing Dremio and SQL, adeptly extracting insights from intraday quote and trade data

Data Science Research Assistant Jan 2024 – Present
Khoury College of Computer Sciences *Boston, MA*

- Curated a dataset consisting of 699 randomized controlled trials (RCTs) by meticulously annotating publications, with the objective of extracting numerical clinical findings crucial for meta-analysis
- Evaluated the performance of instruction-tuned LLMs in automatically extracting numerical data for meta-analysis across various models, including GPT-4, Mistral, Gemma, OLMo, and PMC LLaMA

Data Engineering Intern Jun 2022 – Aug 2022
Space CAMP *Colorado Springs, CO*

- Implemented data streams, state tables, and a JSON API in Apache Kafka with KSQL to deliver precise satellite and mission status updates and facilitate interaction between Python and Kafka for developers

PROJECTS

Author Attribution: NLP | *PyTorch, Python, HuggingFace, Scikit-learn* In Progress

- Developed a comprehensive author attribution project in Python, utilizing Doc2Vec and BERT models from Hugging Face for document feature extraction to discern unique linguistic patterns and writing styles
- Utilized classifiers, such as Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks, to evaluate the effectiveness of embedding techniques and models in accurately attributing authorship

DaveML | *C++, Linear Algebra, Python* Nov 2022

- Implemented ETL module and various regression techniques (linear, ridge, and logistic) using C++ and principles of Linear Algebra and Calculus, such as QR Factorization, Gradient Descent, and cost functions
- Enclosed functionality into a Python module using PyBind11 to enable users with easy experimentation

Hemorrhage Classification with CNN | *TensorFlow, Python, Scikit-learn* Oct 2022

- Utilized TensorFlow to design and implement Convolutional Neural Networks (CNNs) for the classification of brain hemorrhages in CT images, achieving a 70% accuracy rate