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, SQL, C++, 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*

- Constructed a Python-based FIX Protocol message translator deployed on Airflow, enabling the utilization of existing low-touch analytics capabilities for numerous high-touch clients
- Engineered two dynamic Plotly dashboards to enhance the interpretability of quantitative trading signals and facilitate the analysis of order routing, utilizing OneTick and FINRA OTC Transparency Data
- Produced monthly market reports by aggregating terabytes of intraday quote and trade data with SQL

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

- Evaluated the performance of instruction-tuned LLMs in automatically extracting numerical data for meta-analysis, achieving completeness rates up to 82% across various proprietary and open-source models
- Curated a dataset consisting of 699 randomized controlled trials (RCTs) by meticulously annotating research publications, with the objective of extracting numerical clinical findings crucial for meta-analysis

Director of Technical Workshops Sep 2022 – Apr 2023
Disrupt: The FinTech Initiative at Northeastern *Boston, MA*

- Developed and presented four comprehensive Python lectures, covering topics from introductory programming to data manipulation and financial analysis with over 200 students in attendance

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 | *Python, PyTorch, HuggingFace, Scikit-learn* In Progress

- Utilized Doc2Vec and BERT models for document feature extraction to discern unique linguistic patterns
- Implemented classifiers, such as Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks to compare the effectiveness of embedding techniques and model performance on author attribution

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 | *Python, TensorFlow, 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