




Efthimios Vlahos

☎ 516-606-0942 | @vlahos89@gmail.com |  LinkedIn |  GitHub |  Portfolio | 📍 East Meadow, New York

EDUCATION

SUNY Stony Brook University

Master of Science in Applied Mathematics & Statistics, Concentration in Data Science

GPA: 3.52/4.00

- Relevant Coursework: Machine Learning, Big Data Analysis, Data Structures and Algorithms, Linear Programming, Probability Theory, Numerical Analysis

CUNY Hunter College

Bachelor of Science in Mathematics & Physics, Minor in Computer Science

GPA: 3.7/4.0

- Relevant Coursework: Intermediate Mechanics, Quantum Mechanics, Linear Algebra, Abstract Algebra, Vector Calculus, Mathematical Analysis I & II

Stony Brook, NY

Aug 2021 – May 2023

Manhattan, NY

Aug 2016 – Jun 2020

WORK EXPERIENCE

Microsoft

Data Analyst

New York, NY

Nov 2023 – Present

- Developed an **AI Plugin for M365 Chat** using **Python, Azure Machine Learning, and RAG architecture**, achieving a **70% reduction in HR support queries** through automated responses.
- Optimized **hallucination detection frameworks** across multiple Microsoft AI projects, streamlining model validation processes and increasing system reliability by over **35%** by integrating custom metrics and real-time monitoring tools.
- Developed **Power BI dashboards** to visualize the performance metrics of hallucination detection algorithms within AI-driven projects, **enabling real-time monitoring** to enhance accuracy and reliability of content generation.
- Designed **data pipelines using Azure Data Factory, Databricks, and Apache Spark**, enhancing batch processing efficiency by **50%** and **improving the scalability of RAG model training**.
- Worked closely with various teams to align AI strategies with business goals, **enhancing overall team performance through effective communication, presentations, and regular project status updates**.

KPMG

Data Analyst Intern

New York, NY

Jun 2023 – Oct 2023

- Enhanced financial market analysis models using **Python and SQL** for linear regression, achieving a **35% improvement in client portfolio performance** through strategic data-driven optimizations.
- Extracted key insights from economic reports using **NLP techniques** with Python libraries like **NLTK and spaCy**, which increased client engagement by **20%** by refining investment strategies.
- Streamlined the machine learning model training process, implementing automated hyperparameter tuning with **Scikit-learn, MLflow, and TensorFlow**, which cut deployment times by **50%** and enhanced client service efficiency.

PROJECTS

Retail Sales Forecasting with XGBoost | [GitHub](#)

(Python, XGBoost, AWS SageMaker)

- Deployed an **XGBoost model on AWS SageMaker**, achieving an RMSE of 4202.395 and R-squared of 0.965, indicating a **97% variance explanation in weekly sales forecasting through hyperparameter optimization**.

Coupon Optimization Strategy | [GitHub](#)

(Python, SQL, Scikit-Learn, XGBoost)

- Refined coupon distribution strategy using Python and XGBoost, doubling conversion rates from **4.92% to 9.97%**.

NewsletterGen Crew with GUI | [GitHub](#)

(Python, Streamlit, crewAI, LangChain, EXA API, Data Science)

- Developed a **multi-agent AI system using crewAI to automate weekly newsletter generation** with a Streamlit GUI, integrating **EXA API for real-time web scraping** and content analysis.

Chest Cancer Detection | [GitHub](#)

(Python, FastAPI, MLflow, DagsHub, Docker, MLOps, AWS)

- Implemented a CNN-based chest cancer detection platform, achieving **95% recall** with **MLflow** and **DagsHub** integration, deployed via **Docker** on **AWS** for scalable, real-time analysis.

E-commerce Data Analytics Platform on AWS | [GitHub](#)

(AWS, Hive, Spark, Sqoop, SQL, Bash)

- Engineered and deployed a **scalable Hive analytics platform on AWS using Docker**, achieving data ingestion and transformation with **Sqoop, Hive, and Spark** to enhance data-driven decision-making for e-commerce analytics.

TECHNICAL SKILLS

Programming Languages: Python, R, MATLAB, C++, Rust, SQL, Java, Bash

Frameworks & Tools: TensorFlow, Keras, PyTorch, JAX, Scikit-Learn, NLTK, SpaCy, Hugging Face, Pandas, NumPy, PySpark, SciPy, Statsmodels, Langchain, Tableau, Power BI, Looker

Cloud Computing & Big Data: AWS (SageMaker, Lambda, EC2, S3, Redshift), GCP (BigQuery, Vertex AI), Azure (Machine Learning, Data Factory, Kubernetes Service, DevOps), Databricks, Apache Spark, Snowflake, PostgreSQL, MySQL

DevOps & MLOps: Docker, Kubernetes, TFX, MLflow, Kubeflow, Airflow, Terraform, Jenkins, Argo CD, Ansible, GitHub Actions, Prometheus, Grafana, CI/CD, FastAPI, Flask, DVC