

# EFTHIMIOS VLAHOS

(516) 606-0942 | [EfthimiosPortfolio.com](https://efthimiosportfolio.com) | [vlahos89@gmail.com](mailto:vlahos89@gmail.com) | [linkedin.com/in/efthimios](https://linkedin.com/in/efthimios) | [github.com/EfthimiosVlahos](https://github.com/EfthimiosVlahos)

## PROFESSIONAL EXPERIENCE

### CORNERSTONE BUILDING BRANDS

East Meadow, NY [Remote]

Data Scientist

05/2024 – Present

- Engineered an **end-to-end machine learning pipeline** in **Python** (using **scikit-learn**, **NumPy**, **Pandas**) that transformed raw sales data into targeted product recommendations, resulting in a **20% boost** in customer conversion rates and generating an estimated **\$3.5M** in incremental annual revenue.
- Delivered a **production-grade web application** using **Flask** and **Docker** that allowed leadership to input product configurations and obtain real-time predictions on **quote-to-close outcomes** via **hyperparameter-tuned logistic regression** and **random forest models**, driving a **20% increase** in conversion rates.
- Developed scalable **ETL pipelines** in **Databricks** with **Python**, **SQL**, and **Bash** to extract and transform configuration metrics from **Azure Data Lake**, reducing pipeline runtimes by **35%** and lowering excess inventory costs by **15%**.
- Applied advanced **unsupervised learning** techniques (**K-Means clustering** and **PCA**) to segment high-dimensional sales data, enabling the identification of **top-performing product configurations** for optimized production planning.
- Integrated **SQL-based Power BI dashboards** with real-time data feeds, empowering executives and plant managers to make **data-driven decisions** that align production outputs with dynamic market demand.

### MICROSOFT

East Meadow, NY [Remote]

Data Analyst

10/2020 – 05/2024

- Analyzed enterprise-scale datasets with **Python**, **SQL**, and **R** to support an **LLM-based chatbot** solution, achieving a **30% reduction** in hallucination rates and a **25% boost** in user satisfaction through rigorous **anomaly detection** and data cleansing.
- Spearheaded **RAG** projects by fine-tuning **Phi model** using **TensorFlow** and **PyTorch**, resulting in enhanced **contextual accuracy** and improved revenue outcomes for key clients.
- Implemented iterative **A/B testing** and **hyperparameter tuning** with **scikit-learn**, decreasing manual escalations by **20%** and optimizing model performance across diverse client knowledge bases.
- Developed comprehensive **data visualizations** with **Seaborn** and **Matplotlib**, uncovering actionable features that drove targeted improvements in LLM functionality and overall operational performance.

### KPMG

Manhattan, NY [Remote]

Data Analyst Intern

06/2020 – 10/2020

- Streamlined data ingestion and model deployment using **Google Cloud Vertex AI**, **SQL**, and **Python**, reducing training time by **50%** and optimizing portfolio risk analysis for institutional clients.
- Conducted **exploratory data analysis** and built **linear regression models** with **Pandas**, **NumPy**, and **SciPy** to quantify the impact of **interest rates**, **inflation**, and **GDP growth** on asset class returns.
- Delivered data-backed recommendations by integrating **machine learning techniques** with **cloud-based analytics** on **GCP**, reinforcing KPMG's advisory role and optimizing financial market analytics under volatile conditions.

## SKILLS

<b>Programming:</b>	Python, SQL, Bash, C, C++, Java, R, Rust
<b>Data Science:</b>	TensorFlow, PyTorch, JAX, scikit-learn, PySpark, NumPy, Pandas, LangChain, Transformers, Deep Learning, Supervised & Unsupervised Learning
<b>Cloud:</b>	AWS, GCP, Databricks, Snowflake, MySQL, MongoDB, Data Pipeline Automation, Apache Spark
<b>DevOps &amp; MLOps:</b>	Docker, Kubernetes, Terraform, Jenkins, GitHub Actions, Airflow, CI/CD, MLflow, DVC, Grafana, Azure ML, Vertex AI, AWS SageMaker

## PROJECTS

**Generative Deep Learning for Alzheimer's Disease Drug Discovery:** Designed and implemented a generative deep learning pipeline using LSTM and GRU networks in TensorFlow/Keras to process 300K generic and 1.5K targeted chemical compounds from ChEMBL, achieving 94.2% training accuracy and generating 18 novel molecular structures (see GitHub: [Generative-AD-DrugDiscovery](#)).

**Chest Cancer Classification:** Architected a CNN-based classification pipeline in Python that integrates MLflow for experiment tracking and DVC for data version control, streamlining the model development lifecycle and enhancing reproducibility across training iterations (see GitHub: [Chest-Cancer-Classification](#)).

## EDUCATION

<b>SUNY Stony Brook University:</b> Stony Brook, NY	<b>GPA:</b> 3.65 / 4.00
<b>Master of Science: Applied Mathematics &amp; Statistics</b>	<b>Graduation:</b> August 2021 - May 2023
<b>Coursework:</b> Data Structures & Algorithms, Data Science, Machine Learning, Numerical Analysis I, II, and III, Big Data Analytics, Probability Theory, Stochastic Calculus, Linear Programming, Real Analysis	
<b>CUNY Hunter College:</b> Manhattan, NY	<b>GPA:</b> 3.8 / 4.00
<b>Bachelor of Science: Mathematics &amp; Physics (Minor: Computer Science)</b>	<b>Graduated:</b> August 2016 - May 2020
<b>Coursework:</b> Linear Algebra, Vector Calculus, Ordinary Differential Equations, Partial Differential Equations, Probability & Statistics, Quantum Mechanics, Classical Mechanics, Complex Analysis	

## CERTIFICATIONS

[AWS Certified Cloud Practitioner](#), AWS (Apr 24)  
[Large Language Model Operations \(LLMOps\)](#), Coursera (Jun 24)  
[MLOps | Machine Learning Operations](#), Coursera (May 24)  
[Deep Learning Specialization](#), Coursera (Aug 23)