EFTHIMIOS VLAHOS – DATA SCIENTIST

vlahos89@gmail.com | (516)-606-0942 | linkedin.com/in/efthimios-vlahos |github.com/EfthimiosVlahos | East Meadow, NY 11554

PROFFESIONAL SUMMARY

Currently a Data Analyst at Microsoft, I am a dedicated professional with an MSc in Applied Mathematics and Statistics, specializing in Data Science. As an aspiring ML Engineer, I bring over 3 years of experience in data science, adept in utilizing SQL, Python, and R for data analysis, visualization, and machine learning workflows to effectively transform complex datasets into business growth insights.

SKILLS

- **Programming, Libraries, & Cloud Computing:** Python(TensorFlow, Keras, PyTorch, Scikit-Learn, mlflow, Pandas, NumPy, NLTK, Trax, JAX), R, C++, Bash, SQL(PostgreSQL, MySQL, BigQuery), AWS(SageMaker, EC2, S3, Lambda), GCP(Compute Engine, Vertex AI, Auto ML, DataProc, DataFlow), Apache(Spark, Hadoop, Beam, Hive, Airflow)
- **Machine Learning:** Supervised & Unsupervised Learning, Deep Learning(CNNs, RNNs, GRUs, LSTM, Transformers), NLP, Predictive Modeling, Time Series Analysis, Feature Engineering, Model Evaluation, Model Tuning, Validation Methods, Optimization Techniques, Hyperparameter Tuning
- **DevOps, Data Visualization & MLOps:** Docker, Kubernetes, MLflow, Kubeflow, DagsHub, Mage, Tableau, Power BI, Looker, Git, GitHub

PROJECTS/PORTFOLIO

Chest Cancer Detection (Python, TensorFlow, MLflow, DagsHub, Docker, Flask, AWS, MLOps)   New York, NY
Link: Chest Cancer Detection Project   Nov 2023 - Dec 2023
- **Integrated MLflow and DagsHub to manage the machine learning lifecycle,** enhancing experiment tracking and model performance evaluation, resulting in a **20% boost in model accuracy and a 30% reduction in development cycle time.**
- Spearheaded a deployment strategy, **utilizing Docker for containerization and AWS for cloud hosting**, resulting in a scalable web service.
- **Leveraged the VGG16 CNN model for the classification of chest cancer** from CT scan images, achieving an 85% accuracy rate.

E-Commerce Customer Satisfaction (Python, Pandas, Scikit-Learn, XGBoost, ZenML, MLflow, Optuna, Streamlit)   New York, NY
Link: E-commerce Project   Aug 2023 - Nov 2023
- **Successfully integrated MLflow with ZenML, enhancing the machine learning lifecycle from data ingestion to model deployment**, resulting in increased robustness and scalability of models in real-world applications.
- **Enhanced RandomForest model performance by 15% in predictive accuracy**, as measured by comprehensive testing, by implementing Optuna-based hyperparameter tuning and refining model parameters.

News Research Tool Development (Python, Streamlit, Langchain, OpenAI, NLP)   New York, NY
Link: LLM Project   Jan 2023 - Feb 2023
- **Implemented advanced NLP techniques and leveraged OpenAI embeddings along with FAISS vector stores to create a dynamic system** for efficient document splitting, embedding generation, and indexing, facilitating rapid and precise information retrieval from extensive news articles.
- Engineered a proof-of-concept for efficient information retrieval from article chunks, **accelerating research, streamlining data retrieval, and enhancing decision-making in news analysis**.

NYC TLC Predictive Modeling (Python, SQL, GCP, BigQuery, Mage, Tableau)   New York, NY
Link: TLC Capstone Project   Jun 2022 - Jul 2022
- **Designed a data processing architecture using GCP and BigQuery**, effectively streamlining workflows with the Mage Data Pipeline Tool.
- **Applied advanced regression techniques to achieve a 92% adjusted R^2 score on test set** for predicting taxi fare amounts.
- **Conducted strategic A/B testing to reveal the impact of payment types on fares** which led to valuable insights into fare dynamics.

WORK EXPERIENCE

Data Analyst   New York, NY
Microsoft   Nov 2023 - Present
- **Led the development and implementation of NLP techniques to detect hallucinations in AI-generated text**, enhancing the accuracy and reliability of machine-generated information.
- Addressed complex challenges in AI-generated content, **strategizing to improve response accuracy and overall quality.**
- **Communicated key findings and strategies for hallucination detection to stakeholders**, collaborating with data analysts for continuous improvement.

Data Analyst Intern   New York, NY
KPMG   Jun 2023 - Oct 2023
- Advised clients on customer targeting by **assessing data quality and analyzing demographics, leading to targeted outreach**.
- Analyzed data to **target high-value customers based on demographics and attributes**.
- **Developed dashboards to effectively communicate visual findings and insights**. Assessed data quality and completeness in preparation for analysis.

EDUCATION

SUNY STONY BROOK UNIVERSITY, College of Engineering and Applied Sciences   Stony Brook, NY
Masters in Applied Mathematics & Statistics, Data Science Concentration   May 2023
- Cumulative GPA: 3.5/4.0; Dean's List 2022-2023
- **Relevant Coursework**: Machine Learning, Big Data Analysis, Data Structures and Algorithms, Linear Programming, Probability Theory, Numerical Analysis

CUNY HUNTER COLLEGE   Manhattan, NY
Bachelor of Science in Mathematics & Physics, Minor in Computer Science   June 2020
- Cumulative GPA: 3.7/4.0; Dean's List 2016-2019
- **Relevant Coursework**: Intermediate Mechanics, Quantum Mechanics, Linear Algebra, Abstract Algebra, Vector Calculus, Mathematical Analysis I & II