ADITYA SANJAY MHASKE

admhaske@iu.edu | Linkedin/adityamhaske | adityamhaske.com | Github/adityamhaske | Open to Relocation

SKILLS

Languages	Python (PySpark, PyTorch, OpenCV), R, SQL (MySQL, PostgreSQL), Matlab, C++, HTML, CSS
Tools	PowerBI, Tableau, Excel, Azure, AWS, Data Bricks, ETL, GraphQL, Google Analytics, Neo4J, SAS, NoSQL, Hadoop
MLOps	Docker, FastAPI, Flask, Snowflake, REST, Hugging Face, LLM, SAP, Airflow, Spark, GenAI, TensorFlow, MongoDB

EXPERIENCE

Data Scientist Aug 2024 - Present
BMR Infotek Inc. Dublin, CA, USA

- Built an **ML pipeline** with regression algorithms to forecast revenue and customer churn using 3.5M data points and RFM (Recency, Frequency, Monetary) model. Increased accuracy by 34% through optimized feature selection and hyperparameter tuning.
- Extracted, profiled, and analyzed data from diverse sources using SQL, and created operational and financial dashboards. Designed performance monitor dashboards, raising reporting efficiency by 25% and facilitating improved decision making processes.
- Conducted Python correlation analysis to develop data-driven strategies for evaluating marketing campaigns, increasing ROI by 20%.

Data Scientist Dec 2022 - May 2024
Kelley School of Business Bloomington, IN, USA

- Implemented **Llama LLM** with **RAGs** to analyze Customer Satisfaction and Brand Equity for large-size US companies, enhancing predictive accuracy of binary classification by 15% and providing deeper insights into the financial impact on the market.
- Applied **BERT NLP** algorithm to a dataset with over **100 million** rows, achieving 91% testing accuracy in classifying political campaign topics and sentiments, enabling understanding of data-driven and strategic campaign planning.
- Deployed a Cross-Classified Multilevel Model to predict employee performance, for decision-making and improving efficiency.
- Executed comprehensive data analysis to identify patterns and relationships, resulting in a 35% improvement in system performance.

Data Scientist (Full Stack) Intern Sep 2023 - Dec 2023
Twin Cities Innovation Alliance Minneapolis, MN, USA

- Designed and implemented **A/B testing** and **Hypothesis testing** using SQL for data preprocessing, along with statistical modeling like segmentation and regression analysis to evaluate website optimization efforts, resulting in a 17% increase in user engagement.
- Enhanced the web app recommendation system by combining collaborative and content-based filtering methods, utilizing behavioral data and user preferences, leading to a 20% increase in user engagement acquisition.
- Improved client website performance analysis with Python to process and analyze data from over 100,000 user interactions, discovering actionable insights that increased conversion rates by 15%, and providing data-driven recommendations to inform strategic decisions.

Data Engineer May 2021 - Aug 2022
Moonplexus - Healthcare Pune, MH, India

- Led the AWS team to implement **AWS EC2 and Amazon RDS** for efficient database management, establishing an ETL pipeline to compile images, ensuring data integrity, and optimizing data storage for efficient retrieval and Analysis.
- Streamlined the Data Science pipeline and increased scalability by deploying production ready Machine Learning model with **FastAPI, Docker, and Flask**, improving user accessibility by 30% and ensuring robust performance and reliability.
- Automated data retrieval and task execution through SQL integration, reducing execution time by 40%, improving system efficiency.
- Developed ML models using TensorFlow and CNN to achieve 89% accuracy in skin lesion classification, for early diagnosis.

PROJECTS

Stock Prediction on Deutsche Borse using AWS | AWS, Time Series Analysis, S3 Github Project

- Forecasted EUR currency start prices using Python to construct ARCH and GARCH models, for financial decision-making.
- Employed AWS EMR and Sagemaker to process a dataset of over 1000 CSV files, 45.5 million data points stored in an S3 bucket.

Multimodal for Depression Analysis | Python, NLP, Feature Engineering Github Project

- Developed a multimodal system for early-stage depression detection with 90% accuracy, integrating EEG, speech, and facial features.
- IEEE Publication: EEG Based Depression Detection using Ensemble Approach, Signal Processing and Deep Learning

EDUCATION

Master of Science in Data Science, Indiana University Bloomington Aug 2022 - May 2024
Coursework: Artificial Intelligence, Machine Learning, Statistics, Probability, Economics, Data Visualization, Big Data. GPA: 3.6/4.0

Bachelor of Technology in Computer Science, MIT World Peace University Jul 2018 - Jun 2022
Coursework: Computer Networks, Data Modeling, Database Management, Design and Analysis of Algorithms. GPA: 3.7/4.0
Certifications: Business Analyst - Operations Analyst, Customer Analyst, People Analyst (Wharton), AWS, IBM Data Science