VISHAL BAKSHI

Arlington, Virginia | (571) 579-0298 | Mail | LinkedIn | GitHub | LeetCode |

EDUCATION

MS, Data Science, The George Washington University — Washington, DC

B.Tech. Electronics and Communication Engineering, JNTUH — Hyderabad, India

TECHNICAL SKILLS

R, Python, SQL, AWS (EC2, S3, SageMaker, Quicksight), NumPy, SciPy, Pandas, Matplotlib, Seaborn, Plotly, Tableau, MySQL, MongoDB, Apache Spark, Dockers, OpenCV, Beautiful Soup, PySpark, Tensorflow, PyTorch, ScikitLearn, Linux, PyQt5, Git, GitHub

RELEVANT WORK EXPERIENCE

HBots Robotics: Machine Learning Engineer — **Jun 2022 - Aug 2023**

- Developed and deployed an object detection and classification model using MobileNet V1 in PyTorch for real-time product recognition, volume estimation, and inventory tracking, enhancing object detection accuracy by 35%.
- **Led a cross-functional ML project**, integrating **computer vision-based entity resolution** for autonomous robotics applications, improving product scanning efficiency.

The UnEarth: Associate Research & Development Intern (Non-Profit) — **Jan 2021 - Jan 2022**

- Developed an automated ETL pipeline to preprocess student performance data, reducing data cleaning time by 60%.
- Built a predictive model using MLFlow and SageMaker, achieving 98% adjusted R squared accuracy in student performance analysis
- Deployed interactive BI dashboards in AWS Quicksight, providing stakeholders with actionable insights.
- **Conducted workshops for 200+ underprivileged students**, enhancing their skills in basics machine learning, and robotics.

ACADEMIC PROJECTS

Bank Churn Analysis using AWS: [Python, AWS Sagemaker, S3, Quicksight]: **Link**

- Developed a Bank Churn Analysis model leveraging **AWS SageMaker's computation power** and inbuilt ML models, achieving a 98% average score across key metrics through robust preprocessing, RFE feature selection, and RandomForest.
- Utilized **AWS QuickSight** for advanced data visualization, enabling clear insights into customer attrition patterns and actionable trends, empowering stakeholders to make data-driven decisions.
- Streamlined the entire workflow on AWS, leveraging SageMaker for scalable model training and **S3 buckets** for efficient data storage and retrieval, ensuring seamless integration and deployment.

Deep Fake Video Classification: [AWS EC2, Tensorflow , Computer Vision, Python]: **Link**

- Achieved a 50% F1 score in deepfake detection by implementing **CNN + GRU** architecture, leveraging **AWS EC2** instances for scalable model training and storage. Improving performance by 25% through fine-tuning and data augmentation.
- Enhanced real-video classification accuracy by 40% on the CelebDF V2 dataset using class weights, data augmentation, and implementing Error Level Analysis (ELA) to achieve better model generalization.

New York Airbnb Review Analysis: [AWS EC2, NLTK, Spacy, LLM's, Scikit-Learn, PyTorch]: **Link**

- Developed and fine-tuned a **BERT-based sentiment analysis model** with an F1 score of over 90%, analyzing 17,444 Airbnb reviews to classify sentiment with high accuracy and provide actionable insights into guest satisfaction.
- Engineered an advanced NLP pipeline utilizing **T5 for summarization**, generating concise summaries from large text datasets; streamlined review analysis for faster interpretation of key themes like host interaction, cleanliness, and location.

K-means Clustering on Spotify Data: [Hadoop, PySpark, and SQL]: **Link**

- Implemented big data infrastructure on a multi-node cluster using Apache Hadoop and PySpark, building ETL pipelines with Spark SQL to clean and organize over 20 million records for analytics.
- Performed SQL analytics on Spotify datasets to uncover trends and insights, and utilized PySpark MLlib to categorize artists through K-means clustering.

ADDITIONAL INFORMATION

- **Organization:** Led Sahara Research & Development Club as Team Lead for a tenure of 3 years.
- **NGO:** Hosted emerging tech workshop with Telangana Gov. & Deloitte, aiding 800+ underprivileged students. **Link**
- **Award:** Best Social Cause of the Year by BITS PILANI **Link**
- Achieved victory in over 20 technical competitions in robotics at esteemed institutions such as IITs and NITs, reflecting dedication and proficiency in these fields. **Link**