ANUSHA GURUPRASAD

Jersey City, NJ | 347-216-2098
anusha.guruprasad14@gmail.com | linkedin.com/in/anusha-guruprasad | github.com/AnushaG-P

SUMMARY

Data Scientist with an MS in Data Science and **4 years of experience** building software and AI-driven solutions across finance, healthcare, and logistics. Expertise in Python, TensorFlow, and cloud platforms, delivering impactful projects like a Generative AI model boosting drug efficacy by 20%. Published researcher in bioinformatics, presenting genomics findings at international conferences. Awarded at Morgan Stanley's hackathon for a statistical model reducing food waste by 40%.

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems — New York City, New York
Master of Science (MS) in Data Science| **GPA: 3.66/4.0** — May 2024

Visvesvaraya Technological University — Karnataka, India
Bachelor of Engineering (BE) in Computer Science & Engineering | **GPA: 4.0/4.0** — August 2020

TECHNICAL SKILLS

Programming Languages: Python, Java, R, MATLAB, C, C++

Machine Learning and Computer Vision: PyTorch, TensorFlow, Keras, Generative AI, OpenCV , SIFT, Text Mining, Sentiment Analysis

Database Management : Neo4j, Hadoop, HBase, GCP, Cassandra, PostgreSQL, MongoDB, Django

Cloud: Amazon Web Services (AWS), Microsoft Azure, AWS CodePipeline, Google Cloud Platform, Databricks

Data Visualization & Software Tools: Tableau, MS Excel, RStudio, Docker, Git, Jenkins, CI/CD Pipeline

Big Data & Logging Tools: ELK Stack (Elasticsearch, Logstash, Kibana), Apache Spark, Apache Kafka

Additional Skills: Feature Engineering, Statistical Analysis, Model Validation, Forecasting, A/B Testing, API Development-Integration

PROFESSIONAL EXPERIENCE

Music Data Engineer (Contract) — Brooklyn, New York
Arco Data Design Inc. — ***February 2025 - Present***

- Designed and maintained **Python**-based data ingestion pipelines using **Pandas** and **PostgreSQL** to integrate and standardize large-scale music catalog datasets, enabling streamlined reporting and analysis for licensing, royalty tracking, and catalog management.
- Automated deployment via **GitHub** Actions, reducing manual intervention by 30% and accelerating pipeline updates.
- Built **Django**-based applications integrated with **PostgreSQL** to automate workflows, optimizing data management processes. Implemented **GitHub CI/CD pipelines** for testing and deployment, ensuring 99% uptime and alignment with comprehensive data integration systems.
- Created **Python** scripts for dynamic reporting solutions and automated **ETL tasks**. Leveraged GitHub for version control and collaboration, improving data accuracy by 25% and harmonizing integrated data systems across teams.

Senior Machine Learning Developer (Part-time) — New York City, New York
Shoptaki

- **Senior Machine Learning Developer (September 2024 – March 2025):**
  - Designed a **time series forecasting** integrated with **XGBoost** to predict structural breaks across various intervals, enhancing **forecasting** accuracy and supporting advanced analytics.
  - Developed Adobe Analytics **dashboards** to visualize trends and predictions, effectively communicating insights to stakeholders and ensuring optimal functionality of data-driven applications.
  - Conducted performance testing and refined models to ensure efficiency and optimization in financial forecasting systems.
- **Machine Learning Developer(June 2023 – September 2023):**
  - **Led a data science team of 10 people** to create a clustering model that improved full investment conversions and developed a regression model for accurate property tax estimation using **TensorFlow**, **PyTorch**, and **Keras**.
  - Identified gaps in investment decision-making and lease analysis tools for emerging real estate investors.
  - Increased platform revenue by 20% and reduced investor decision time by 15%, significantly enhancing investment efficiency.

Graduate Research Assistant (Full-Time) — New York City, New York
Pace University – Research Paper — ***November 2023 – June 2024***

- Led a comprehensive study analyzing **10000+ genomic and metabolomic datasets** to uncover molecular mechanisms behind Rafflesia's parasitic behavior, to understand the host-pathogen dynamics in plant biology.

- Characterized microbiome, metabolome, and transcriptome profiles of Rafflesia-Tetrastigma symbiosis using **R and Python**, creating **interactive Excel and Tableau dashboards** to visualize biotic interactions for research engagement.
- Engineered **R pipelines** to automate analysis of XCMS and LCMS mass-spec data, **processing 10,000+ data points to generate statistical insights and box plots** comparing regional species interactions.
- Co-authored a peer-reviewed paper titled **"**Microbes and Metabolites of Tetrastigma**"** and presented findings at the Annual Plant Biology Conference (ASPB) in Hawaii, reaching 200+ global researchers and earning recognition for methodological innovation.

Associate Application Developer Bangalore, India

Unisys ***September 2020 – July 2022***

- Led a team of **5 developers** as the **DevOps** team representative, mentoring them on product workflows and code management processes. Provided hands-on guidance in backend database handling and **AWS workflows**, ensuring the team was well-equipped to manage and deploy scalable solutions effectively.
- Developed and deployed a **Java**-based travel experience platform hosted on **AWS Cloud**, enhancing customer satisfaction by optimizing workflows for multi-destination ticket creation, exchange, refund, and cancellation.
- Coordinated with cross-functional teams to streamline processes and ensure seamless handling of complex ticketing operations.
- Leveraged **ELK Stack** to implement an application log management system for real-time monitoring and error tracking, improving system reliability and operational insights.
- Automated **CI/CD pipelines** using **Jenkins** and AWS, incorporating shell scripting to streamline routine **ETL tasks** and enhance deployment efficiency.
- Integrated **Bitbucket** for **version control** and **JIRA** for task tracking, ensuring effective collaboration and efficient development cycles within the CI/CD framework.
- **Recognized by management and team leads** for demonstrating **client-centricity, integrity, and curiosity** in delivering innovative solutions.

PROJECTS

Impact of Generative AI on Personalized Drug Treatments for Pre-existing Conditions May 2024

- Developed a **Generative AI model with Random Forest classification** in **python** to predict Adverse Drug Reactions (ADRs) **based on genetic profiles**.
- Achieved **78.49% accuracy in predicting ADR**s and **improved drug treatment efficacy by 20%,** leading to more personalized medicine.
- Predicted a **15% reduction in ADRs**, contributing to potential cost savings and better patient outcomes.

Hand Joint Detection using Computer Vision & MATLAB (View it Live!) January 2023

- Identified the need for accurate hand joint detection for improved gesture recognition and interaction in applications.
- Optimized **Facebook's Detectron2** to achieve 98% accuracy and developed a finger joint identification system using **OpenCV and scikit-learn**, reaching 92% accuracy.
- Deployed on Streamlit for real-time interaction, offering instant feedback on detected objects and enhancing user experience.

Machine Learning-Driven Market Basket Insights with Tableau November 2023

- Addressed the need for improved data visualization and analysis in market basket scenarios to enhance decision-making using **R programming**.
- Developed Tableau dashboards for market basket analysis, integrating **R for statistical modeling using Linear Regression, K-means clustering, Logistic Regression, Support Vector Machines, and Random Forest**.
- Achieved a 25% improvement in visualization clarity and a 30% increase in model accuracy, enhancing overall decision-making efficiency.

ACHIEVEMENTS

Won the "A Byte of Food", Morgan Stanley Hackathon – "Code to Give" | Atlanta Community Food Bank

- Contributed as a backend developer using **NodeJS and NoSQL MongoDB**, developed a statistical modeling solution that increased community outreach by 32% and ensured equitable food distribution, resulting in a 40% reduction in food waste.

Publications

- Co-authored research paper titled **"Microbes and Metabolites of Tetrastigma: Deciphering the Ecology of Host Choice of the Plant Parasite and World's Largest Flower, Rafflesia (Rafflesiaceae)"** in a reputed publication. (Read Here)
- Co-authoring a paper titled "**Metagenomic and Metabolomic Insights into Microbial and Chemical Drivers of Rafflesia's Parasitic Life Cycle**", which is currently under review.