KASI MAJJI

kasi.majji24@gmail.com
Chicago, IL, USA
LinkedIn: linkedin.com/in/kasi-majji/
GitHub: https://github.com/kasimajji

SUMMARY

Data Scientist with 2.5+ years of experience in designing and optimizing scalable data pipelines and machine learning solutions across cloud platforms like AWS, GCP and Azure. Proven expertise in building end-to-end ML workflows, integrating real-time data processing, and deploying models in production environments. Adept at leveraging tools such as Apache Spark, Kafka, and Airflow to enhance data processing efficiency. Recently graduated with a master's in information systems, with a focus on advanced analytics and cloud technologies.

EDUCATION

Northern Illinois University - Master's, Information Systems. **August 2023 – May 2025**

Coursework: GPA: 3.4

- Business Analytics SAS & Advanced Predictive Data Analytics, Business Data Networks & Cybersecurity, Business Application Database Mgmt. System, Business System Analysis & Design.

EXPERIENCE

Tata Consultancy Services - Data Scientist, India **May 2022 - July 2023**

- Collaborated with cross-functional teams (underwriters, actuaries, and claims analysts) to define business objectives and translate them into data science solutions.
- Developed predictive models using Python (scikit-learn, XGBoost) for:
  - **Risk scoring** during underwriting using Logistic Regression algorithm.
  - **Claim fraud detection** which reduced fraudulent claims by 17%.
  - **Customer churn prediction** for policy renewals which improved customer retention rate by 12% using targeted renewal campaigns powered by these models.
- Utilized **Pandas**, **NumPy**, and **Seaborn** for exploratory data analysis and hypothesis testing to uncover trends in customer behavior, premium defaults, and claim frequency.
- Engineered and selected features from structured data in AWS Redshift and semi-structured data stored in **S3** using **PySpark** and **Glue**.
- Deployed models using **AWS SageMaker**, integrated into existing CI/CD workflows (via **GitHub Actions** and **Terraform**) inherited from previous engineering experience.
- Created interactive **Power BI** and **Tableau** dashboards to visualize key insurance KPIs across lifecycle stages: quote, bind, policy issuance, claim initiation, and renewal.
- Authored and optimized **SQL queries** to join policy, claims, customer, and agent data from Redshift for input to ML pipelines.

Tata Consultancy Services – Data Engineer**,** India **Jan 2021 - May 2022**

- Led the design and optimization of **cloud-based Data Lakes and Warehouses implemented for Beneficiary details** using AWS (S3, Redshift, Glue, Lambda), reducing data retrieval time and storage costs by 15**%.**
- Architected **real-time data ingestion pipelines** using **Kafka and Spark Streaming**, significantly reducing data processing latency.
- Designed **CI/CD pipelines** for data workflows using **GitHub Actions and Terraform**, improving insurance product deployment automation by 15**%**.
- Mentored **junior engineers**, conducting code reviews and training sessions, which improved team productivity

PROJECTS

End to End ML Pipeline for Truck Delay Classification | GitHub

- Developed ML system to classify truck delays using **AWS Sage Maker Pipelines**, automating training, evaluation, and deployment in a **CI/CD setup**.
- Led the full ML lifecycle from data ingestion and **preprocessing** to **monitoring**, **incorporating data drift** detection and **retraining logic** to ensure sustained model performance in production.
- Leveraged **Python**, **Scikit-learn**, **Hops Works** feature store, **Weights and Baises** for Experimentation tracking, Docker, **Evidently AI** for monitoring and evaluating ML models and data in production, and **AWS** (Lambda, Kinesis, CDK, S3) to build a scalable, end-to-end cloud-native pipeline.

Flower Image Classification using ResNet50 | End-to-End Deep Learning Project | GitHub

- Built a deep learning application flower classification model using **ResNet50** with **transfer learning** to classify 5 flower types, achieving high accuracy on real-world image data.
- Implemented a comprehensive machine learning lifecycle encompassing **data preprocessing**, **model training**, **evaluation**, and **deployment**, utilizing **ML flow** for experiment tracking and model versioning.
- Tech Stack: Python, PyTorch, ResNet50, ML flow, Streamlit, Pandas, Matplotlib.

CERTIFICATIONS

- Microsoft-certified Azure AZ900, Cloud Administrator AZ-104.
- Certified for completing the Java course by IIT Bombay.

SKILLS

Programming & Frameworks: Python, R, SQL, PySpark, scikit-learn, XGBoost, TensorFlow, Keras, PyTorch.
Data Engineering & Big Data: Apache Spark, Apache Airflow, AWS (S3, Lambda, Redshift, Glue, SageMaker, Azure, GCP, Databricks.
Databases & Storage: MySQL, Snowflake, MongoDB, AWS Redshift, S3.
ML Lifecycle & MLOps: MLflow, Weights & Biases, Docker, Terraform, GitHub Actions, Evidently AI.
Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Power BI, Tableau, Excel.
Tools & Platforms: Git, GitHub, Eclipse, Google Collab, MS Office.
Soft Skills: Cross-functional Collaboration, Communication, Analytical Thinking, Problem Solving, Time & Project Management, Technical Documentation, Stakeholder Engagement.