RASHMITHA PAGADALA

Jersey City, NJ | +1 (626) -313-8262 | rashmithapagadala34@gmail.com| LinkedIn

PROFESSIONAL SUMMARY

Results-driven Data Science and Machine Learning professional with extensive experience in developing predictive models, building scalable data pipelines, and delivering data-driven insights to solve complex business challenges. Proficient in leveraging Python, SQL, PySpark, and machine learning frameworks like TensorFlow, Scikit-learn, and XGBoost to build end-to-end solutions for data analytics and forecasting. Skilled in designing and deploying ETL pipelines, automating data workflows with tools such as Apache Airflow, AWS, and Google Cloud Platform, ensuring seamless data integration and analysis. Adept at using Power BI, Tableau, and other visualization tools to transform complex data into actionable insights for decision-making. Strong ability to collaborate with cross-functional teams to optimize data processes, enhance data-driven strategies, and improve operational efficiency. Proven track record of driving 15% customer retention improvement and enhancing operational processes through predictive analytics and automation.

TECHNICAL SKILLS

Programming & Scripting: Python, SQL, R, Java, C++, Bash, JavaScript, HTML, CSS, TypeScript
Machine Learning & AI: Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, LightGBM, CatBoost, Transformers, BERT, LLaMA, spaCy, OpenCV, YOLOv5, RNN, CNN, AutoML, Hugging Face
Deep Learning & NLP: LSTM, GPT, T5, Sentence Transformers, Word2Vec, NLTK, TextBlob
Data Engineering & Pipelines: PySpark, Apache Spark, Apache Beam, Apache Airflow, dbt, AWS Glue, Talend, Apache NiFi, Kafka, Flink, ETL/ELT
Cloud Platforms: AWS (S3, EC2, Lambda, Redshift, RDS, SageMaker, Athena, Glue), GCP (BigQuery, Vertex AI, Dataflow, Pub/Sub), Azure (ADF, Synapse, Databricks, Blob Storage)
Databases & Warehousing: PostgreSQL, MySQL, MongoDB, Cassandra, DynamoDB, Snowflake, BigQuery, Redshift, SQL Server
Visualization & Business Intelligence: Power BI, Tableau, Looker Studio, Plotly Dash, Streamlit, Matplotlib, Seaborn, Altair, Excel
MLOps & DevOps: MLflow, DVC, Docker, Kubernetes, Git, GitHub Actions, Jenkins, Terraform, Prefect, Kubeflow, Weights & Biases
Data Analysis & Tools: Pandas, NumPy, SciPy, StatsModels, Jupyter, Google Colab
Project & Workflow Management: Agile, Scrum, JIRA, SDLC, CI/CD, Asana, Trello

WORK EXPERIENCE

Project Intern
National Remote Sensing Centre | Hyderabad, India **Aug 2023 - Feb 2024**

- Developed a Flask-based geospatial imagery platform utilizing REST APIs, streamlining access for 500+ researchers to real-time satellite data, which resulted in a 70% improvement in data retrieval speed, enhancing productivity across research teams.
- By leveraging NumPy and OpenCV, optimized image preprocessing workflows, reducing spatial alignment time by 50% for multi-band satellite imagery, significantly accelerating the analysis for agricultural research and improving overall data quality.
- Engineered a cost-effective, serverless architecture with AWS S3, Lambda, and API Gateway, delivering scalable access to over 2 TB of geo-tagged satellite data, and reducing infrastructure costs by 30% while supporting large-scale environmental analysis.
- Revamped data ingestion pipelines with Celery and Redis, enabling asynchronous processing, which decreased data delivery delays by 40% and improved the timeliness of satellite data for national environmental monitoring programs.
- Built interactive Power BI dashboards to visualize NDVI, cloud cover, and crop cycles, providing real-time insights and facilitating data-driven decision-making, leading to a 20% improvement in reporting efficiency for agriculture-related stakeholders.
- Applied PostGIS on PostgreSQL to enhance geospatial querying performance, improving data retrieval speed by 60% for high-resolution satellite imagery, enabling faster and more accurate analysis of critical environmental datasets.
- Implemented automated data validation using Pandas Profiling and Great Expectations, ensuring schema integrity and eliminating 90% of potential data issues, thus enhancing the accuracy and consistency of the data used for machine learning models.
- Collaborated with cross-functional teams of cloud engineers and scientists to align data pipelines with national mission analytics, streamlining workflows and contributing to the successful deployment of high-impact environmental and agricultural projects.

Project Intern
BDL India | Hyderabad, India **Apr 2023 - Jul 2023**

- Developed a real-time object detection system for missile trajectory tracking using YOLOv5 and OpenCV, which improved simulation review efficiency by 50%, reducing the time required for each flight test and accelerating decision-making processes.
- Built a Kafka-Python alert system that identified real-time trajectory anomalies, cutting response delays by 40%, and enhancing the accuracy of simulations with timely alerts, ensuring faster intervention and safety measures during missile tests.
- Created Plotly Dash dashboards to visualize missile trajectory data, including bounding boxes, prediction scores, and detection metrics, enabling real-time test cycle analysis, improving collaboration, and ensuring test results were easily accessible to all stakeholders.

- Enhanced model performance by integrating TensorRT and ONNX, improving FPS handling by 30% in fast-motion simulation environments, ensuring better performance for high-speed missile trajectory simulations and reducing lag during testing.
- Integrated Flask with JavaScript and WebSocket to provide seamless real-time visualization of model predictions, resulting in live updates during defense trials and improving the model's adaptability for on-the-fly analysis and control.
- Automated log data processing with pandas and regex, reducing data wrangling time by 60%, speeding up simulation output cleaning, and enhancing the efficiency of the data analysis pipeline for missile test results.
- Utilized MLflow for experiment tracking and model versioning, which ensured reproducibility and consistency across deployments, allowing for efficient and organized management of machine learning models throughout the testing lifecycle.
- Collaborated with trajectory engineers to adjust detection thresholds and model metrics, achieving 20% improvement in tracking accuracy, optimizing the alignment of model outputs with flight path data, and increasing simulation reliability.

PROJECTS

Customer Purchase Behavior Analysis

- Segmented customers using RFM-based analysis on SQL-sourced transaction data, applying K-Means clustering to identify customer cohorts, improving targeted marketing strategies by tailoring offers to high-value segments.
- Trained a Random Forest classifier to predict customer purchase probability, integrating results into Power BI dashboards for real-time insights, leading to more informed business decisions.
- Improved customer retention by 15% through personalized campaigns based on predictive behavioral insights, optimizing marketing spend and increasing ROI by focusing efforts on the most promising customer groups.

Crop Recommendation System (ML Based)

- Developed a dual-model recommendation system using Random Forest and XGBoost to forecast the best crops for specific regions, utilizing data on soil, rainfall, and temperature to provide accurate recommendations.
- Enhanced prediction accuracy by 20% through advanced feature engineering, incorporating soil nutrient ratios and microclimatic indicators, resulting in more precise crop predictions for farmers.
- Integrated model predictions into Power BI dashboards, delivering actionable insights to farmers, improving crop selection reliability and boosting agricultural productivity across seasons.

Billing & Logistics Data Automation

- Built an ETL pipeline using Apache Airflow, Python, and SQL to aggregate data from multiple departments and automate report generation, streamlining business operations and saving time for the finance and logistics teams.
- Centralized data into Amazon Redshift, standardizing schemas and applying data transformations to ensure improved accuracy in business reports and consistent data analysis across departments.
- Developed Power BI dashboards for KPI monitoring, cutting manual report cycles by 30% and providing quicker, actionable insights for decision-makers in logistics and finance operations.

EDUCATION

Master of Science in Data Science **Sep 2024 – May 2026**
Montclair State University, Montclair, NJ | GPA: 3.9/4.0

Bachelor of Science in Data Science **Jun 2020 – May 2024**
CMR Engineering College, JNTUH, Hyderabad, India | GPA: 3.5/4.0

CERTIFICATIONS

• Google Data Analytics Professional Certificate – **Coursera**
• IBM Machine Learning with Python – **Coursera**
• Data Engineering Fundamentals – **LinkedIn Learning**
• Deep Learning Specialization – **Coursera**
• Data Science for Business Leaders – **DataCamp**
• Python for Data Engineering – **IBM SkillsBuild**

PUBLICATIONS

• **Classification of Stars and Galaxies using ML** – Published research utilizing CNN and Random Forest on SDSS astronomical datasets for improved classification.
• **IoT-based Automated Agriculture System** – Integrated machine learning algorithms with sensor data for precision farming, contributing to smarter agriculture practices.