

PRUDHVI MARPINA

Email: prudhvi.marpina2@gmail.com
LinkedIn: www.linkedin.com/in/prudhvi-marpina

Phone: +1 (425) 463-5402
GitHub: https://github.com/prudhvi-marpina

SUMMARY

Data and machine learning professional with four years of experience delivering end-to-end analytics and machine learning solutions in fast-paced startup environments. Proficient in Python, SQL, and cloud platforms including AWS, Azure, and GCP. Experienced in building scalable data pipelines, automating ETL workflows, and deploying machine learning models in production. Developed dashboards and reports to support business teams and collaborated across functions to translate data into actionable insights. Focused on building reliable systems that connect data, software, and business value.

PROFESSIONAL EXPERIENCE

Data Scientist | Tech Friday | Hyderabad, India

June 2021 - July 2023

- Built and deployed a serverless data ingestion system using Python, AWS Lambda, and S3 to unify API and database sources with greater than 95% reliability, enabling product analytics and ML pipelines.
- Designed and implemented a preprocessing pipeline using PySpark to clean, validate, and transform data in BigQuery, improving input consistency and accelerating data readiness for ML models by 25%.
- Developed and applied ML models (logistic regression for churn prediction, k-means for segmentation) using Scikit-learn, to identify high-risk users and personalize outreach, boosting engagement by 15 to 20%.
- Automated ML retraining workflows with Airflow DAGs, adding schema versioning, retry logic, and alerts, reducing pipeline failures by 30%.
- Created interactive dashboards in Power BI and Looker Studio using SQL to visualize model performance and user behavior, supporting decisions that increased user engagement by 12%.
- Collaborated with product and marketing teams to interpret model results and build audience profiles, enabling data-driven features and targeted campaigns.

Junior Data Analyst | Ahlada Engineers Ltd | Hyderabad, India

June 2019 - May 2021

- Contributed to a company-wide analytics initiative to streamline operations and supply chain reporting, enabling integrated data workflows across production, inventory, and finance departments.
- Explored sales and receivables data using Python and SQL to uncover patterns in delayed payments, informing credit policy adjustments that improved the average collection cycle by 15 percent.
- Constructed Tableau dashboards using SQL and Excel to surface key metrics on production efficiency and inventory flow, reducing manual reporting time by 60 percent.
- Processed and cleaned order and delivery data exported from SAP ERP to support supply chain reporting and planning, contributing to a 10 to 15 percent improvement in on-time fulfillment during quarterly review cycles.
- Partnered with supply chain and finance teams to validate data accuracy and align reporting across 10+ KPIs, ensuring consistency in monthly business performance reviews.

TECHNICAL PROJECTS

UTDIGI - Chatbot for UT Dallas

- Built a modular Scrapy - Playwright pipeline to crawl the entire utdallas.edu domain, extracting high-quality content and storing 3,000+ structured, cleaned documents in MongoDB with retry and validation.
- Transformed scraped content using spaCy NER and Mistral-7B summarization with custom chunking to reduce input size by ~60% while preserving key entities and context.
- Implemented a RAG pipeline with LangChain, Sentence Transformers, ChromaDB, CrossEncoder reranking, and LLaMA-2 via Hugging Face, improving top-3 answer relevance by ~25% in internal tests.
- Launched a FastAPI backend and Streamlit interface with Redis caching and source-linked responses for real-time, multi-turn query handling under 8 seconds.

Optimizing Gardein's Market Strategy: Data-Driven Insights for Growth

- Built a predictive model using Scikit-learn to evaluate market performance across 8 U.S. regions, identifying over \$10M in projected revenue opportunities.
- Analyzed 3 years of sales data using Pandas and NumPy to uncover product flavor and packaging insights enabling up to 25% uplift in targeted regions.
- Applied K-Means clustering and time-series analysis to identify seasonal demand patterns, improving forecast accuracy by 15% in test scenarios.
- Created Power BI and Looker dashboards to visualize trends, reducing insight turnaround time by 30% in stakeholder reviews.

EDUCATION

The University of Texas at Dallas | Richardson, Texas

August 2023 - May 2025

Master's in Data Science and Artificial Intelligence

GPA: 3.97 / 4.0

Birla Institute of Technology and Science, Pilani | Goa, India

Bachelor's in Electrical and Electronics Engineering

SKILLS

- **Programming Languages & Data Science Tools:** Python, SQL, R, C, C++, Java, JavaScript TypeScript, Data Structures, Algorithms, Pandas, NumPy, spaCy, Matplotlib, Microsoft Fabric, Microsoft Power Platform, Data Management, EDA, Data Wrangling, Feature Engineering, Data Modeling, Model Evaluation, Data Analysis, Econometrics, Business Analytics.
- **Machine Learning, Deep Learning and Generative AI:** Scikit-learn, TensorFlow, PyTorch, Keras, Regression, Classification, Ensemble Methods, Hyperparameter Tuning, Causal Inference, Cross Validation, Time Series Analysis, A/B Testing, NLP, Hugging Face, LangChain, LLMs, RAG, Agentic AI, Prompt Engineering, Fine tuning, GANs, Diffusion Models, MLOps, Product Analytics, Statistics, AI/ML Workflows.
- **Cloud Computing, Big Data and Databases:** Apache Spark, PySpark, Airflow, Databricks, Hadoop, Hive, dbt, Delta Lake, Kafka, Spark Streaming, AWS, Azure, GCP, Snowflake, BigQuery, Redshift, PostgreSQL, MySQL, MongoDB, SQLite, Cosmos DB, VectorDB, Parquet, Columnar Storage, ETL/ELT, Stream and Batch Processing, Query Optimization, Kubernetes, Data Engineering, Data Architecture, Data Integration, Data Warehousing.
- **Business Intelligence, Tools and Development Frameworks:** Power BI, Tableau, Looker Studio, Excel, VBA, Git, CI/CD, Data Visualization, Jenkins, Docker, REST APIs, FastAPI, Streamlit, JIRA, Agile, Scrum, Client Engagement

CERTIFICATIONS

- **Applied LLM and AI Agent Engineering Specialization:** Gained expertise in advanced AI, specializing in LLMs, RAG, and Agentic AI. Built solutions using Google Gemini API and LangChain for robust RAG with embeddings and vector stores. Skilled in creating AI agents with function calling, LangGraph, prompt engineering, fine-tuning, and guardrails for reliable results.
- **AI Data Engineering Specialization:** Gained hands-on experience in the data engineering lifecycle-including ingestion, storage, transformation, and serving-using Apache Spark, Airflow, Kafka, dbt, and AWS to build scalable data systems.

CLUBS AND ORGANIZATIONS

- **Envision (Data Visualization Club), UT Dallas, Member:** Participated in data hackathons, visualization challenges, and technical events focused on Power BI, Tableau, and storytelling with data.
- **AWS Club, UT Dallas, Member:** Participated in cloud hackathons, workshops, and certification events focused on AWS technologies and career development.
- **Nirmaan, BITS Pilani, Volunteer Member:** Contributed to education and community development initiatives as part of a BITS Pilani alumni led NGO focused on inclusive and sustainable social impact.

AWARDS

- Lars Magnus Ericsson Fellowship in Management