

# PRUDHVI MARPINA

**Email:** marpina.prudhvi@gmail.com

**LinkedIn:** www.linkedin.com/in/prudhvi-marpina

**Phone:** +1 (425) 463-5402

**Website:** https://prudhvi-marpina.github.io/portfolio/

## SUMMARY

Data Scientist and Machine Learning Engineer with 4+ years of experience designing and deploying ML systems, recommendation models, and data pipelines across cloud platforms (AWS, GCP, Azure). Skilled in Python, SQL, PySpark, A/B testing, distributed systems, and model deployment. Driven to apply ML and data science to build scalable solutions that enhance decision-making, elevate user experiences, and deliver measurable business impact.

## SKILLS

**Programming and Software Engineering:** Python, SQL, Java, C++, R, FastAPI, React, Node.js, REST APIs, GitHub, GitHub Actions, CI/CD, Data Structures, Algorithms, Object Oriented Programming, Microservices Architecture

**Machine Learning and AI:** PyTorch, TensorFlow, Scikit-learn, XGBoost, spaCy, Hugging Face, LLMs, RAG, Generative AI, Model Deployment, Recommendation and Ranking Systems, Hyperparameter Tuning, MLflow

**Data Science and Analytics:** Statistics, Statistical Modeling, Exploratory Data Analysis (EDA), Feature Engineering, Experimentation, A/B Testing, Causal Inference, Metrics Design, Time Series Forecasting, Pandas, NumPy, Matplotlib, Tableau, Power BI

**Data Engineering and Cloud:** Apache Spark, PySpark, Kafka, Airflow, Hadoop, Hive, ETL/ELT, Data Warehousing, Databricks, Snowflake, Query Optimization, AWS (SageMaker, Lambda, S3, Redshift), GCP (BigQuery), Azure, Redis, Docker, Kubernetes, Terraform, Distributed Systems, MLOps Pipelines

**Certifications:** DeepLearning.AI Data Engineering Professional Certificate (AWS), SnowPro Associate – Platform (Snowflake)

## PROFESSIONAL EXPERIENCE

**Software Engineer (Volunteer) | One Community | California, USA**

August 2025 – Present

- Built a React front-end and Python API integrated with OpenAI GPT to automate reporting summaries, reducing manual reporting time by 25%.
- Developed a Python dashboard (Plotly/Dash) to visualize 5,000 MongoDB time entries, showing keyword frequency and volunteer participation trends for managers.
- Conducted A/B tests on automated post scheduling pipelines, validating a 12% CTR lift and guiding rollout of cron-based posting strategy.
- Implemented DistilBERT for text classification and spaCy NER for entity extraction on volunteer reports and outreach posts, reducing manual tagging by 30%.

**Data Scientist | Tech Friday | Hyderabad, India**

June 2021 – July 2023

- Designed serverless ingestion pipelines in Python with AWS Lambda and S3 to capture transaction data from APIs and databases with >95% reliability.
- Engineered PySpark workflows to clean, join, and aggregate millions of transactions into feature tables in BigQuery, cutting preparation time by 25%.
- Deployed Scikit-learn models (logistic regression for churn, k-means for segmentation) as REST APIs integrated into the platform for real-time predictions.
- Automated retraining in Airflow with DAGs for feature refresh and model updates, with job monitoring that reduced pipeline failures by 30%.
- Conducted A/B tests on retention campaigns powered by churn models and visualized results in Power BI, demonstrating an 8% uplift in user retention.

**Data Analyst | Ahlada Engineers Ltd | Hyderabad, India**

June 2019 – May 2021

- Streamlined ERP data extraction by scripting Python (Pandas) and SQL workflows on Tally ERP, reducing monthly report prep time by 30%.
- Automated receivables and delivery reporting with scheduled Python jobs, reducing manual processing effort by 40%.
- Developed Tableau dashboards on SQL tables to track plant output, material usage, and order fulfillment, cutting reporting cycles from 5 to 2 days.
- Validated ERP data with supply chain and finance teams, ensuring consistency across 10+ KPIs used in monthly reviews.

## TECHNICAL PROJECTS

**UT Dallas Digital Assistant – RAG-based Chatbot**

May 2025

- Built a RAG-based chatbot indexing 3,000+ documents in MongoDB and ChromaDB; applied NER, summarization, and reranking, improving top 3 answer accuracy by 25% and reducing response time to <8s with Redis caching and a Streamlit user interface for real-time scalability.

## EDUCATION

**The University of Texas at Dallas | Richardson, Texas**

August 2023 – May 2025

*Master of Science in Data Science and Artificial Intelligence (GPA: 3.97 / 4.0)*

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

*Bachelor of Engineering in Electrical and Electronics Engineering*