Prudhvi Raj

prudhvimikey@gmail.com | 347-871-8879 | LinkedIn | Github | Portfolio | United States

PROFESSIONAL SUMMARY

Results-driven MS in Computer Science graduate with hands-on experience developing machine learning models, automating data pipelines, and extracting business insights from complex datasets. Proficient in Python, SQL, TensorFlow, and Scikit-learn, with a strong foundation in statistical analysis and deep learning. Completed a 6-month internship where I improved forecasting accuracy by 23% and reduced reporting delays by 40% across real estate and tech data environments. Passionate about building scalable ML systems that drive data-informed decision-making.

TECHNICAL SKILLS

Programming & Libraries: Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, XGBoost), SQL, Git, Streamlit, Jupyter Notebook
Machine Learning & AI: Supervised & Unsupervised Learning, Time-Series Modeling (LSTM), Ensemble Learning, NLP, Cross-Validation, Hyperparameter Tuning, Model Evaluation, Bias Mitigation, LLM Integration (Mistral, LLaMA, MythoMax via OpenRouter)
Feature Engineering & ML Ops: Temporal Features, Lag Variables, Frequency Encodings, ETL Automation, Schema Validation, Data Versioning, Experiment Tracking, Model Reproducibility, Pipeline Automation
Data Analysis & Visualization: EDA, Outlier Detection, Statistical Profiling, Tableau, Power BI, Matplotlib, Seaborn
Databases & Cloud Platforms: MySQL, PostgreSQL, AWS S3, Google Colab, SQLite, Apache Spark
Math & Statistics: Hypothesis Testing, Probability, Linear Algebra, Optimization, Regression Analysis, PCA, Statistical Inference
Collaboration & Workflows: Agile Sprints, GitHub Workflows, Technical Documentation, End-to-End ML Ownership, Cross-functional Collaboration, Data Handoff to Engineering

PROFESSIONAL EXPERIENCE

Data Science Intern | Company: Vimtra Ventures Private Limited *January 2023 – July 2023*

- **Improved real estate price forecasting accuracy by 23%** across regional datasets by building and fine-tuning ML models using Scikit-learn, TensorFlow, and PyTorch; integrated ensemble learning with time-based and categorical feature encodings.
- **Engineered 20+ production-grade features** (e.g., temporal lags, seasonality flags, frequency encodings), increasing model interpretability and reducing retraining cycles.
- **Processed and validated 1.2M+ transaction records** using Pandas and NumPy, ensuring 100% schema integrity and preventing downstream model failures in staging environments.
- **Built real-time Tableau dashboards** to track revenue, pricing trends, and model performance, used weekly by strategy and investment teams across 3 business units, **reduced reporting delays by 40%**.
- **Owned end-to-end ML pipeline** - from data sourcing and EDA to model deployment and stakeholder reporting; documented codebase and experimentation logs to support handoff to engineering team.

PROJECTS

Customer Churn & LTV Intelligence Platform | SQL, ETL Automation, Python, LSTM, Random Forest, NLP, OpenRouter API
Built a churn prediction system with 91% accuracy and a GenAI chatbot using LSTM, Random Forest, and OpenRouter LLMs.

- **Achieved 91% churn classification accuracy** and stabilized LTV forecasts by leading a **3-member team** to build a sequence-aware prediction system using **LSTM** for behavioral data and **Random Forest** for tabular features.
- **Reduced ETL runtime by 50%** across **100K+ customer records** by automating ingestion with 14 modular **SQL scripts** and a **Python-based loader**, saving **8+ hours per data refresh cycle**.
- Developed **Muffin**, a GenAI assistant using **OpenRouter LLMs** (Mistral, LLaMA, MythoMax) to answer business-facing questions like *"Why is this customer likely to churn?"* with over **90% relevance accuracy**.
- Combined **sentence-transformer embeddings** with fallback chaining to generate **100+ plain-language insights** in a real-time **Streamlit chatbot**, improving model explainability for business users.

Real Estate Price Forecasting App - Connecticut | Python (Scikit-learn, XGBoost, TensorFlow, Matplotlib, Seaborn, Joblib)
Developed a real-time housing price predictor using ensemble models and feature engineering in Streamlit.

- **Improved housing price prediction accuracy by 18% (R^2 = 0.82)** by training ensemble models on statewide datasets enriched with **20+ engineered features** (e.g., lag variables, seasonal encodings, locality segmentation).
- Deployed a real-time **Streamlit dashboard** for user-driven forecasts, **reducing manual lookup time by 90%** and increasing accessibility for non-technical stakeholders.
- Applied advanced **feature engineering** techniques to enhance signal clarity and generalization, boosting model robustness across geographies and property types.
- Presented results as part of coursework, incorporated feedback to refine **UI/UX, error handling,** and input validation logic.

EDUCATION

Pace University | New York, United States *September 2023 – May 2025*
Master of Science in Computer Science | Grade: 3.7/4.0