SANDEEP UNDURTHI

Logan, UT 84321
sandy.undurthi@gmail.com
linkedin.com/in/sandeep-undurthi — github.com/sandeepundurthi

Summary

Recent Master's graduate in Computer Science with hands-on experience in machine learning, data analysis, and developing end-to-end data-driven solutions. Skilled in Python (Pandas, Scikit-learn, TensorFlow, PyTorch), SQL, Spark, and MLOps tools such as MLflow and Vertex AI. Experienced in LLM fine-tuning, prompt engineering, RAG pipelines, and causal modeling. Passionate about applying data to solve real-world problems. Seeking an entry-level role as a Data Scientist, Machine Learning Engineer,Data Engineer or Data Analyst.

Education

Master of Computer Science, Utah State University, Logan, UT
GPA: 3.8/4.0
Relevant Coursework: Machine Learning, Data Mining, Neural Networks, Advanced Algorithms, Database Systems, Social Network Analysis

Technical Skills

Languages & Tools: Python (Pandas, NumPy, Matplotlib, Scikit-learn, TensorFlow, PyTorch, Hugging Face), SQL, Git, VS Code, MySQL, Spark, Streamlit
ML/DS Methods: LLM Fine-tuning, Prompt Engineering, RAG, Multi-Agent AI, Supervised & Unsupervised Learning, Causal Inference, A/B Testing, Uplift Modeling, LTV Modeling, GenAI (BERT)
MLOps & Deployment: MLflow, Vertex AI, Airflow, Ray, LangGraph, Logging, Model Monitoring
Other: Data Extraction, Entity Recognition, Schema Mapping, Cloud Deployment

Projects

UN Country Profiles Dashboard *Technologies: Streamlit, Plotly, Pandas, Git, GitHub, Streamlit Cloud*
Developed an interactive data dashboard to explore and compare development indicators (e.g., GDP, population, internet usage) for 200+ countries using UN datasets. Built country-wise profile viewer, top-10 visualizations, and comparison module. Implemented global choropleth map and CSV export. Deployed on Streamlit Cloud.

Smart Scheduling: Patient No-Show Predictor *Technologies: Python, Streamlit, Scikit-learn, SHAP, SMOTE, Pandas, Matplotlib*
Built a pipeline to predict patient no-shows (100K+ hospital records). Engineered features like lead time, appointment day, SMS reminders. Achieved 76% accuracy with Random Forest; improved recall by 95% using SMOTE. SHAP explanations and overbooking simulator integrated into Streamlit app.

House Price Prediction System *Technologies: Python, Pandas, Scikit-learn, XGBoost, Lasso/Ridge Regression*
Developed a regression model using the Ames Housing dataset. Handled missing values, categorical encoding, and scaling. Applied Linear, Ridge, Lasso, and XGBoost models. Tuned hyperparameters and documented outputs in Jupyter notebook.

Amazon Review Fraud Detector with SHAP Explainability *Technologies: Python, Streamlit, Scikit-learn, SHAP, TextBlob*
Built a web app to detect potentially fake or paid reviews. Engineered features like sentiment polarity, subjectivity, and mismatch. Used Isolation Forest for anomaly detection with SHAP visualizations. Users can analyze reviews for trustworthiness.

Restaurant Review Sentiment Analysis *Technologies: Python, Scikit-learn, Pandas, TF-IDF, Gradio, Matplotlib*
Classified restaurant reviews as Positive, Neutral, or Negative using TF-IDF bigrams and supervised models. Compared NB, LR, SVM, RF achieving 96% with SVM. Deployed model with Gradio interface and visualized model performance.

Twitter Sentiment Analysis Engine *Technologies: Python, Scikit-learn, NLTK, Pandas, Matplotlib*
Built pipeline to classify Twitter sentiments. Included tokenization, lemmatization, and TF-IDF vectorization. Evaluated multiple models using accuracy, precision, recall, and F1-score with Streamlit deployment.

Fake Job Posting Detection *Technologies: Python, NLP, Scikit-learn, Streamlit, TF-IDF, Logistic Regression*
Built an NLP classifier for job post fraud detection using TF-IDF and Logistic Regression. Addressed imbalance with SMOTE. Included word cloud visuals and deployed real-time prediction web app with confidence scoring.

Certifications & Activities

Certifications: Machine Learning Specialization (Coursera), Python for Data Science (DataCamp)
Simulations: British Airways Data Science Simulation, Deloitte Australia Analytics Simulation (Forage)
Community: Machine Learning Study Group Organizer, Kaggle Competition Participant