Krishna Mallik Nanduri

knanduri.m@gmail.com • linkedin.com/in/nanduri-k • github.com/kmallik60 • Portfolio

SKILLS

- **Programming**: Python, SQL, Scala, Java, PySpark, R, C, C++, JavaScript
- **Libraries**: Pandas, NumPy, Scikit-learn, Seaborn, Matplotlib, NLTK, ggplot, dplyr, gensim, spaCy, OpenCV
- **Software/Tools**: Tableau, Power BI, Docker, Git, Unix, Airflow, Jenkins, Microsoft Office, JMP, SAS
- **Database Systems**: AWS RDS, MySQL, IBM DB2, OracleDB, MongoDB, Azure SQL Server
- **Big Data Technologies**: Hadoop, Hive, Apache Kafka, Apache Spark, MapReduce, Spark Streaming
- **Cloud Technologies**: GCP (Dataproc, BigQuery), AWS (EC2, RDS, SageMaker, Kinesis, S3), Azure (SQL Server, Data Lake, Databricks)

WORK EXPERIENCE

Christian Tyler Properties LLC **Dec 2023 – Present**

Data Scientist ***Tampa, USA***

- Engineered an investor support chatbot using **LLMs (GPT-4o)** on **Azure Databricks** with **Spark**, trained on internal data repositories; enabled scalable data processing and reducing turnaround time by 85%, with secure deployment via **Docker** and **Azure Virtual Machines**
- Architected and migrated a comprehensive **financial transaction** database for **600+ investors** from Excel to **Azure SQL Server**; designed a robust relational model with **SSIS automation** to ensure **data integrity** and facilitate **financial reporting**
- Implemented an end-to-end ETL process leveraging **Power Apps**, **Power Automate** and **OpenAI GPT-4o** to ingest real time lead data, automate **ActiveCampaign CRM** Deal creation, and generate **AI-Driven** note summaries reducing manual processing time by **96%**
- Designed and deployed an **AWS-based (EMR, Lambda, S3) pipeline** to analyze **1000+ monthly** investment leads, centralizing A/B-tested campaign data in **Amazon Redshift**; developed **Power BI dashboards** to optimize investment marketing and client acquisition
- Engineered a real-time data pipeline on **AWS (EC2, Kinesis, S3)** and **Azure SQL Server** for unified EB-5 investment analytics from 25,000+ client interactions, **reducing** data latency to **<10 seconds** and accelerating investment inquiry conversion by **30%**
- Designed **Power BI** dashboards using **DAX** and **SQL** for comprehensive **EB-5** and **US Real Estate Fund** performance analysis, including interest/return calculations, revenue/expense forecasting, and cash flow insights, driving over **$20M in savings**
- Developed **SQL Server** queries (joins, window functions, CTEs, Stored Procedures) to streamline investor/project data for **compliance audits**; reduced discrepancies by 90% and delivered 15+ **ad-hoc Power BI dashboards**, boosting reporting efficiency by 35%

Via Separations **May 2022 – Dec 2022**

Machine Learning Engineer ***Boston, USA***

- Implemented **MLops** practices by developing a powerful ML pipeline on **AWS Sagemaker**, employing **TCN**, and **LSTM** for deep learning and **Time Series Analysis** (**ARIMA**, **SARIMA**) on the sensor data stream, resulting in a 60% accuracy boost for product maintenance forecasts
- Spearheaded SQL database migration and setup using **AWS RDS and EC2**, designing optimized queries, stored procedures, and views to streamline data retrieval. Integrated database with a front-end interface developed in **Retool** and **JavaScript**, achieving a 40% reduction in data analysis time.
- Employed **SAS JMP** for statistical analysis to optimize data normalization for membrane form factor experiments, facilitating precise cross-comparisons and improving experimentation efficiency, resulting in a 20% reduction in analysis time
- Enhanced data integrity by 40% using automated purchase order processing via **GCP APIs** and data ingestion optimization with **Python-based Bash scripts**. Deployed **Airflow** for **CI/CD orchestration** and **Jenkins** with **Git** for streamlined pipeline management

Bluebonnet Data **Jan 2022 – May 2022**

Data Analyst ***Minneapolis, USA***

- Engineered automated data pipelines using **Databricks** to extract, transform and load (ETL) USA census data with Minneapolis' 159 precincts voter data for voter behavior analysis, resulted in 70% reduction in data processing time
- Developed **Tableau** dashboards integrating trends and patterns from behavior analysis, providing stakeholders with real-time insights, improving the accuracy of decision-making by 30%, streamlining campaign strategies
- Conducted geospatial analysis using **ArcGIS**, **QGIS**, and **Geopandas** to visualize voter sentiment across precincts and applied **K-means clustering** to identify behavior groups, resulting in optimized outreach methods and a 20% increase in voter engagement effectiveness.

Verzeo **Jan 2021 – Jun 2021**

Data Scientist ***Hyderabad, India***

- Utilized a diverse range of machine learning and deep learning models, including the **Random Forest Regressor**, **XGBoost** for ensemble learning, and **LSTM** as a recurrent neural network (RNN), for accurate time series forecasting
- Optimized big data workflows using **Apache Spark**, **MapReduce**, and **Hadoop**, cutting processing time by 65% over Python-based methods

Projects

Click-stream Data Analysis for E-commerce Personalization **"Real-Time Processing, Personalized Recommendations"**

- Developed a real-time ETL pipeline with **Scala**, **Spark**, and **Kafka** to process clickstream data in **BigQuery**. Built and implemented recommendation models, increased user engagement by 20% and increased conversion rates by 15%

Energy Consumption Prediction **"Boosting Techniques, Feature Engineering"**

- Used **XGBoost**, **LGBM**, and **CATBoost** ensemble models with feature engineering, Bayesian Optimization, and Halving Grid Search to predict HVAC energy consumption from a 40M-record dataset, achieving 1.27 RMSLE

EDUCATION

Northeastern University, Boston, MA **Sept 2021 – Aug 2023**

Master of Science in Data Science *CGPA: 3.8/4.0*

Gandhi Institute of Technology and Management University, India **Jun 2017 – Jun 2021**

Bachelor of Technology in Computer Science and Engineering *CGPA: 3.75/4.0*