Krishna Mallik Nanduri

(617) 595-8107 • knanduri.m@gmail.com • linkedin.com/in/nanduri-k • github.com/kmallik60 • krishnananduri.com

PROFESSIONAL SUMMARY

- Data Scientist with 3+ years expertise in analytics, predictive modeling, and scalable data engineering, providing actionable insights and data-driven strategies for various business domains
- Independent Data Science lead at Christian Tyler Properties LLC, delivering end-to-end solutions across data engineering, statistical modeling, visualization, and automation, contributing to significant operational efficiencies
- Proficient in Python and PySpark, with hands-on experience in large language models (LLMs) like GPT-4, as well as regression, predictive modeling, clustering, and time-series forecasting to solve complex business challenges
- Experienced in SQL and data warehousing, including migrating legacy data systems (e.g., Excel-based workflows) to structured databases like SQL Server, enabling scalable analytics and improved data integrity
- Strong knowledge of cloud and big data platforms, including Azure, AWS, and GCP, with experience deploying solutions on EC2, Databricks, SQL Server, and BigQuery to ensure secure, scalable operations
- Skilled in developing dashboards using Power BI, Tableau, and Excel, effectively communicating insights and supporting strategic decision-making
- Experienced in implementing data governance practices, ensuring data accuracy, compliance, and security across analytics and reporting systems

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**
- Migrated a large scale legacy investor database, including financial records for 600+ investors, from Excel based to **Azure SQL Server**; designed a **relational model** and used **SSIS** and custom automation scripts to secure stakeholder-only access
- 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%**
- Orchestrated a scalable lead engagement system using **GCP Dataproc** and **GCP Cloud Functions** to capture and analyze engagement for **1000+ monthly leads**. Centralized **A/B-tested campaign insights** in **BigQuery** and visualized performance data in **Looker Studio**, optimizing marketing strategies in collaboration with the **marketing team**
- Engineered a real-time WhatsApp messaging pipeline with a **Flask webhook on AWS EC2**, ingesting and processing engagement data from **25,000+ contacts** via **AWS Kinesis Data Streams** and **AWS S3** feeding an **EC2 consumer** into **Azure SQL Server** for unified analytics, cutting latency to **under 10 seconds** and empowering sales to convert leads **30% faster**
- Leveraged **Power BI**, **DAX**, **SQL**, and **Paginated Reports** to build dynamic dashboards automating revenue & expense tracking and payment schedules for investment portfolios, providing real-time cash flow insights and saving over **$2M**

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
- Employed **Apache Spark**, along with **MapReduce** and **Hadoop**, to streamline big data processing and conduct exploratory data analysis (EDA), achieving a significant 65% decrease in processing time compared to traditional Python 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 deployed recommendation models, increasing user engagement by 20% and boosting conversion rates by 15%

IoT Sensor Data Pipeline for Predictive Maintenance **"Predictive Maintenance, Streaming Analytics"**

- Built a real-time **IoT** data pipeline with **Scala** and **Spark Streaming** on **Azure Databricks** for predictive maintenance, enabling anomaly detection and reducing equipment downtime by 40%. Processed real-time sensor data with **Kafka**, storing results in **Azure SQL Database** for analysis. Optimized maintenance schedules, reducing repair costs

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*