

Krishna Mallik Nanduri

📍 Bridgewater, NJ (Open to Relocation) 📞 (617)-595-8107 ✉ nanduri.k@northeastern.edu **in** [nanduri-k](#)  [kmalik60](#)

 [Portfolio](#)

EDUCATION

- **Northeastern University, Boston, MA** Sept 2021 – Aug 2023
Master of Science in Data Science CGPA: 3.8/4.0
- **Coursework:** Supervised Machine Learning, Unsupervised Machine Learning and Data Mining, Intro to Data Management, Data Management for Analytics, Parallel Processing for Data Analytics, Algorithms.
- **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



SKILLS

Programming: Python, SQL, Pyspark, R, C, C++, Java, JavaScript
Libraries: Pandas, NumPy, Scikit-learn, Seaborn, Matplotlib, NLTK, ggplot, dplyr, gensim, spaCy, media pipe, OpenCV
Software/Tools: Tableau, PowerBI, Microsoft Office, Docker, Git, Unix, JMP, SAS, AWS (EC2, RDS, SageMaker, Athena)
Database Systems: AWS RDS, MySQL, IBM DB2, OracleDB, MongoDB
Big Data Technologies: Hadoop, Hive, Apache Kafka, Airflow, Apache Spark, MapReduce
Cloud Technologies: GCP (Vertex AI, Tensorflow), AWS, Azure Databricks

WORK EXPERIENCE

- **Via Separations, Watertown, MA USA** May 2022 – Dec 2022
Machine Learning Engineer Co-op “Neural Networks, Predictive Modeling”
 - 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.
 - Collaborated with cross-functional teams to migrate 1 TB of diverse Excel-based data to an optimized **AWS RDS Relational database** on **AWS EC2**, utilizing **MySQL**, resulting in a 35% reduction in product experimentation data retrieval time
 - Spearheaded front-end development with **Retool** and **JavaScript**, achieving a 40% reduction in data analysis time, seamlessly integrated the front-end system with AWS database using **SQL Queries**, stored procedures, views, functions.
 - 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% via automated vendor purchase order processing using **Google Cloud Platform (GCP) APIs**, optimizing data ingestion with **Python-based Bash scripts**. Integrated **Git version control** and **CI/CD pipelines** with **Jenkins**.
 - Developed A/B tests and experiments to assess the effects of varied product testing parameters. Utilized statistical analysis to drive data-informed process improvements in the product performance
- **Bluebonnet Data, Minneapolis, MN, USA** Jan 2022 – May 2022
Data Analyst Intern “Complex Visualizations, Clustering, Geospatial Analysis”
 - Developed **Tableau** and **Power BI** 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
 - 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
 - Utilized **ArcGIS** and **QGIS**, **Geopandas** for geospatial analysis, creating plots to visualize voter sentiment metrics in various precincts. Analyzed voter behavior patterns and preferences, optimizing campaign strategies.
 - Employed **K-means clustering** for cluster analysis to identify five distinct behavior groups, effectively categorizing voter demographics, optimizing outreach methods, resulting in a 20% increase in voter engagement effectiveness
- **Verzeo (Microsoft Authorized Education Partner), Hyderabad, India** Jan 2021 – Jun 2021
Data Science Intern “PySpark, Parallel Processing”
 - 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 LSTM model for a significant 45% reduction in SMAPE, improving microbusiness density forecasting performance
 - 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
 - Implemented feature engineering techniques (lag features, rolling statistics, exponential smoothing) and performed **Principal Component Analysis (PCA)** for dimensionality reduction, resulting in a reduction of data noise by 10%

Projects

- **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.
- **Mental Health Analysis on Social Media Posts | ** “Natural Language Processing”
 - Applied **NLP** techniques and deep learning models to analyze and cluster Reddit posts, predicting suicide risk at 80% accuracy using feature extraction methods such as **TF-IDF** and **Word2Vec**.
- **Healthcare Chatbot Using LLMs on GCP** “Large Language Models, Google Cloud”
 - Developed a healthcare chatbot with LLMs on **GCP**, integrating **dbt** and **GraphQL** for seamless data workflows, ensuring real-time medical guidance and information delivery.
- **Diabetic Retinopathy Identification Project**
 - Enhanced detection with **OpenCV** preprocessing and **CNN** developed in **Keras & TensorFlow**, leveraging deep learning for efficient medical image classification and analysis.