ANUKEERTHI REDDY POTHEPALLI

apothepalli1@student.gsu.edu | (470) 838-5598 | LinkedIn | Portfolio

WORK EXPERIENCE

Data Scientist, Institute for Insight at* Georgia State University** ***Jan 2024 - Current

- Deployed machine learning algorithms, including **Random Forest, LightGBM, and XGBoost**, for **fraud detection** in insurance claims, leading to strengthened fraud prevention through advanced predictive analytics.
- Processed health insurance claims using **NLP techniques** like Named Entity Recognition (NER) to categorize patient names and procedures, enhancing data quality and analysis accuracy by 12%.
- Analyzed 23M Stack Exchange posts using **GPT models (3.5, 4.0)** and **text analytics tools** (BERT, sklearn) enhancing AI question-answer accuracy by 35%.
- Fine-tuned and deployed BERT models using Hugging Face for **text summarization** and **classification**, achieving a 25% improvement in classification accuracy and a 15% increase in ROUGE scores.

Data Scientist at* Kids2 Inc.** ***Aug 2023 – Dec 2023

- Collaborated with product and marketing teams to enhance existing **forecasting models** using time series techniques (Prophet, AutoTS, ARIMA), reducing Mean Absolute Error by 7.5% over baseline models.
- Queried and modeled 100K+ data points with **SQL, Spark, and Power BI**, refining retailer order trends and boosting post-launch success by 8.5% through advanced analytics.
- Implemented machine learning models (**Random Forest, XGBoost, NLTK**) to analyze product reviews, improving market performance by 15% with actionable insights.

Data Scientist at* Tata Consultancy Services (TCS), Bangalore, India** ***Apr 2021 - Jul 2023

- Ran Naive Bayes & logistic regression models for multi-class text classification based on case description. Predicted estimated time for each case resolution optimizing company's resources by **30%** and client's investment.
- Developed a machine learning demand prediction solution on Azure Databricks, improving accuracy by 5% through feature engineering.
- Led a 10% boost in **client satisfaction** and ensured compliance using **AI models**, maintaining zero escalations over 11 months.
- Reduced file processing time by 10% and increased accuracy by 18% using unsupervised **machine learning, Explainable AI – SHAP,** and **text analytics** on medical device data.
- Leveraged advanced SQL queries from Salesforce ECM to **visualize KPIs** and complaint metrics through Power BI and Tableau dashboards, enhancing stakeholder decision-making by 15% with actionable insights.

EDUCATION

Georgia State University, J. Mack Robinson College of Business **Atlanta, GA, United States**

Master's in data science and Analytics (Business Concentration) *Aug 2023 - May 2025*

Courses: Machine Learning, Data Programming, Statistical Foundations, Scalable Analytics, Predictive Analytics, Deep Learning Analytics, Business Intelligence, Data Management, Data Visualization, Text Analytics, Graph Analytics

PROJECTS

ECG Classification with CNN and LSTM: *python, pandas, numpy, sklearn, keras, matplotlib, random forests, SVM, gradient boosting*

- Created and implemented a robust **Neural Network model** (CNN + LSTM) using pytorch for Multi-class ECG classification, allowing real-time monitoring of heart rate patterns; the algorithm processed data from over 10,000 ECG recordings to ensure reliability in clinical settings with an accuracy equaling the **SOTA(99%).**

Enhancing Car Buying Experiences with LLMs*: generativeAI, GPT – 3.5, LLM, LinQ embeddings*

- Engineered and deployed a cutting-edge AI car recommendation system using LinQ **LLM Vector Database** and GPT-3.5/4, enhancing decision-making efficiency by **9x** and cutting costs by 30% through state-of-the-art **generative AI** architecture.

Scalable restaurant recommendation system Optimization*: pyspark, GCP (DataProc), cloud computing resources, looker studio*

- Engineered a scalable recommendation system(PySpark, RDD) utilizing **Google Cloud's** cluster resources, analyzing 5 million+ data points. Built visualizations in Looker Studio, enhancing efficiency, and delivering key insights about restaurant reviews data(yelp).

SKILLS

Competencies: Linear Regression, Classification (Logistic, SVM, Random Forests, k-nearest neighbor), Gradient Boosted Machines, Text Processing, Transformers, SARIMAX, Prophet, LIME, SHAP, Partial Dependence Plots, Parameter estimation, Hypothesis Testing, Word Embedding (Word2Vec, BERT), Statistical Distributions, Large Language Models (LLM), Neural Networks(CNN, RNN, ANN)

Programming Languages: Python, Structured Query Language (SQL), R-Programming, NoSQL

Tools/Frameworks: Power BI, Microsoft Excel, Pandas, Numpy, StatsModels, Matplotlib, Seaborn, PySpark, Sklearn, Scikit-learn TensorFlow, Keras, Tableau, Google Cloud Platform (GCP), Azure Machine Learning Studio, Asana, Git

CERTIFICATIONS & ACHIEVEMENTS

- *Oracle Cloud: Oracle Cloud Infrastructure 2024 Generative AI Professional* *Jul 2024*
- *DataCamp: Data Analyst Associate Certificate* *Jan 2024*
- *Institute of Insight, Robinson College of Business: Best Data Analytics Poster* *Apr 2024*