Charan Teja Kurakula

Fairfax, VA, 22030 **Email:** kcharanteja437@gmail.com

Charan Teja Kurakula | LinkedIn **Phone:** 5714376001

PROFESSIONAL SUMMARY

Masters student in Data Analytics Engineering with a strong interest in using data to solve real-world problems. Skilled in Python, SQL, and R, and familiar with cloud tools like AWS and GCP. Comfortable working with data pipelines, machine learning models, and visualizations to find useful insights. Always eager to learn and contribute to a team that values data-driven thinking and collaboration.

EDUCATION

George Mason University Fairfax, Virginia
Master's in data Analytics Engineering; GPA: 4.00 **January 2024 – Present**
BMS Institute of Technology and Management Bengaluru, India
Bachelor of Technology – Electronics and Communication Engineering; GPA: 8.12 **August 2019 – May 2023**

SKILLS

- **Languages:** Python, SQL, R, C, Java.
- **Frameworks:** Pandas, NumPy, Matplotlib, Scikit-Learn, Matplotlib, Seaborn, Fast API, SciPy, Streamlit, TensorFlow, PyTorch, Retrieval-Augmented Generation (RAG), Large Language Models (LLM's), OpenAI API, SpaCy, NLTK, Hugging Face Transformers.
- **Technologies:** Artificial Intelligence (AI), Machine Learning (ML).
- **Tools:** Excel, Power BI, Tableau, MongoDB, LangChain, Linux, ETL.
- **Platforms:** Visual Studio Code, Colab, Jupyter Notebook, Amazon Web Services (AWS), Google Cloud Platform (GCP), Databricks, R Studio, GitHub.
- **Data Skills:** Data Visualization, Data Analysis, Data Warehousing, Statistics,Data Processing.
- **Soft Skills:** Excellent Communication, Team Collaboration, Research Aptitude.

WORK EXPERIENCE

George Mason University **May 2025 - Present**
Graduate Student Researcher *Fairfax, Virginia*

- Collaborated with Dr. Liao to design and implement a machine learning pipeline for monitoring structural defects in bridges from data in the real world offered by the Federal Highway Administration (FHWA).
- Engineered machine learning pipelines that achieved 74% accuracy in classifying bridge conditions as 'Good,' 'Fair,' and 'Worse' based on defect statistics and Elasticity modulus values.

PROJECTS

PDF RAG Chatbot **March 2025 – May 2025**

- Designed a chatbot system using Fast API and Lang Chain that interprets, and answer questions based on the content of uploaded PDF documents with the help of pdfplumber and GPT-4o vision to extract visual and textual data from complex technical PDFs with reasonable accuracy.
- Embedded the extracted content into Chroma DB using OpenAI's embedding models to allow natural language queries for contextual semantic wellness and generated a scalable and modular API so that both PDF uploads could be managed seamlessly, thus returning understandable, readable sentences via the chatbot interface.

Credit Score Prediction Using Machine Learning **January 2025 – May 2025**

- Created a credit score prediction model using machine learning models like Random Forest, XGBoost with an accuracy of 83% by performing complex preprocessing, including KNN imputation, feature engineering, and Chi-square test for independence of features with the help of TensorFlow.
- Engineered a basic Flask web application for real-time prediction; included SHAP and LIME for model interpretability and fairness analysis. Established competing sources of data and fairness-sensitive approaches to improve accuracy and financial inclusion.

Big Data to Improve Insurance Complaint Handling **October 2024 – December 2024**

- Processed large volumes of insurance complaint data and used PySpark for effective processing and analysis of the data and determined primary concerns like delays in handling complaints and classified them in terms of root causes and resolution strategies.
- Used clustering algorithms to determine patterns in the frequency with which types of complaints arose and the time to resolve them.

CABG Surgery Hospital Performance Analysis **February 2024– April 2024**

- After assessing California CABG operation data, statistical summaries, visualizations, and wrangling of significant health care characteristics were produced using R and Python.
- Large datasets have been examined using AWS S3, Glue Data Brew, and SQL to determine which areas had the lowest performance and to impact data-driven health care improvements. Trends have been identified by comparing hospitals to the state average.

Library Management System **March 2024 – April 2024**

- Outlined a database for a public library network to store books, members, and borrow history and staff information. Designed the ER diagram and then converted the ER diagram into relational schema, including the necessary primary and foreign keys and created the tables using SQL brand in MySQL and then imported real-world data from eight actual CSV files into the database.
- Developed SQL queries that included JOIN, subqueries and aggregations to find overdue books, most-borrowed items and author statistics.
- Provided several useful suggestions, like reminders daily for overdue and automatic suspension/reactivation of members based on borrowing history.

Loan Eligibility Prediction Using Data Analysis **October 2022 – May 2023**

- Collected and analyzed loan application data from banking industry customers to identify customer profiles and assess consequent decisions made based on the characteristics.
- Initiated and implemented a machine learning model by making use of the Python programming language to predict loan approvals based on the features of income level, credit history, amount of loan, and employment history. Assessed model performance about accuracy metrics and tuned hyperparameters, improving quality and stability of predictions for insights.
- Built a Web Framework using Streamlit which tells about the loan status and configured visualizations and dashboards, to facilitate risk indicators and better support data-informed lending decisions for the clients.

CERTIFICATIONS

Google Data Analytics Certificate (Google) **July 2024 – December 2024**

- Acquired experience in data cleaning, visualizing and analyzing data in tools including Excel, SQL and Tableau using the Google Data Analytics Professional Certificate.
- Learned how to process through the data analysis process of ask, prepare, process, analyze, share, and act to solve business problems in the real world with data.
- Completed capstone projects using real world datasets, and practiced exploratory data analysis, dashboards, and communicating insights to non-technical stakeholders.

AWS Academy Cloud Foundations (AWS) **April 2024**

- Acquired knowledge about fundamental cloud concepts, global infrastructure, and foundational services (EC2, S3, RDS, IAM).
- Obtained experience deploying virtual servers, configuring storage services, understanding AWS pricing models and best practices.
- Established core knowledge in cloud computing, networking, security, and architecture by applying industry-recognized best practices recommended for AWS certification.