BHARGAVI KADAMBARI
M:469-351-9574 | bhargavikadambari32@gmail.com | LinkedIn | Github |

PROFESSIONAL SUMMARY

Data Science Master's graduate with expertise in machine learning, data engineering, and big data analytics. Skilled in Python, SQL, Tableau, and cloud platforms (AWS, GCP). Experienced in predictive modeling, ETL pipelines, and real-time analytics, with a strong track record of optimizing business performance through AI-driven insights and automation. Adept at leveraging data-driven strategies to enhance operational performance, risk management, and user experience across diverse industries. Strong analytical and problem-solving skills.

EDUCATION

University Of North Texas, Masters in Data Science, GPA: 4.0/4.0 **Jan 2024 – Present**
Jawaharlal Nehru Technological University, Bachelors in Electronics & Communication**,** GPA: 7.9/10 **July 2017 - Jun 2021**

TECHNICAL SKILLS

- **Data Engineering & Cloud Computing:** NumPy, Pandas, ETL Pipelines, Data Transformation, Data Warehousing**,** GCP, BigQuery, Dataproc, AWS (S3, Glue, Lambda, Redshift)**,** Apache Spark, Apache Airflow, Kafka, Hadoop**,** Open Refine, Apache Hive, Pandas **&** Real-time Data Processing, Streaming Data Pipelines
- **Machine Learning & Data Analytics:** Python, SQL, PySpark, Scikit-learn, TensorFlow, Keras, XGBoost, LSTM, GRU, ARIMA, NLP, Predictive Analytics, Deep Learning, Statistical Modeling & Exploratory Data Analysis (EDA), Feature Engineering
- **Data Visualization & Software Development:** Tableau, Power BI, Matplotlib, Seaborn, Plotly, Git, Jupyter Notebook, Visual Studio Code, IntelliJ, Agile, Scrum, Waterfall Methodologies

CERTIFICATES

Google (GCP) Certified Professional Data Engineer
Oracle Database SQL Certified Associate

ACADEMIC RESEARCH PROJECTS

University Of North Texas **Denton**, TX
Project 1: 09/2024-Present
Data Life Cycle in Homeless Encampment Requests

- **Developed a scalable data processing pipeline** for analysing homeless encampment requests in Dallas using **Google Cloud Platform (GCP), BigQuery, Apache Spark, and Hive**. Optimized **query performance** through **partitioning and clustering**, enhancing **data retrieval speed** and **efficiency**.
- **Executed advanced data cleaning and transformation** using **Open Refine**, ensuring **high-quality, structured datasets** for analysis. Applied **SQL-based queries in BigQuery** to extract insights on **encampment trends, response times, and peak request periods**, enabling **data-driven decision-making** for city authorities.
- **Deployed Apache Spark and Hive on Google Dataproc** to process **large-scale datasets** efficiently. Leveraged **distributed computing** for **high-speed data transformation** and **real-time analytics**, demonstrating expertise in **cloud computing, ETL pipelines, and big data management**.

Project 2: 05/2024-08/2024
Stock Price Forecasting Project

- **Developed a robust stock price forecasting model** leveraging **machine learning algorithms** such as **Random Forest, XGBoost, LSTM, ARIMA, and GRU**, integrating **real-time market data** to enhance prediction accuracy and adapt to dynamic financial trends.
- **Engineered key features** from **historical stock data, technical indicators, and market factors**, optimizing model performance through **hyperparameter tuning, cross-validation, and evaluation metrics (MSE, R-squared)** to deliver precise financial insights.
- **Deployed the model via a scalable cloud-based infrastructure** using **Python (Pandas, NumPy, Scikit-learn, and TensorFlow), Yahoo Finance API, AWS/GCP, and Jupyter Notebook**, facilitating seamless access to forecasts through a **user-friendly web interface**.

Project 3: 03/2024-05/2024
Body Fat Percentage Prediction and Analysis

- **Developed a predictive model for body fat estimation** using **Python (Pandas, NumPy), machine learning techniques, and data visualization tools (Tableau, Matplotlib, Seaborn)** to analyze the impact of factors like metabolism, alcohol consumption, and body measurements.
- **Performed extensive Exploratory Data Analysis (EDA) and feature engineering** on a **252-row dataset from Kaggle**, uncovering correlations between body fat percentage, age, weight, and hip circumference using **scatter plots, correlation heatmaps, bar charts, and bubble charts**.
- **Designed interactive visualizations in Tableau** to present insights on **gender-based fat distribution, alcohol's impact on body fat, and key influencers of fat accumulation**, facilitating data-driven health and fitness decision

PROFESSIONAL EXPERIENCE

GCP Data Engineer at WIPRO **Hyderabad, INDIA**
Project 1: **12/2022-12/2023**
TELSTRA-Pharmacy Analytics and Optimization System

- Developed a data-driven solution to enhance **operational efficiency** and **improve patient outcomes** in pharmacy workflows.
- Designed, developed, and implemented high-performance **ETL pipelines** using the **Python API (PySpark)** of **Apache Spark**.
- Worked on processing **unstructured data** in **JSON** format and converting it to **structured data** in **Parquet** format by performing several transformations using **PySpark.**
- Optimized data processing **workflows** using **Python** and **SQL**, **reducing costs** by **15%** and improving overall **system performance**.
- Real-time **Tableau dashboards** were developed to present **KPIs**, enhancing **efficiency by 25%.**
- Conducted a **deep analysis** of the **SQL execution plan** and recommended hints, **restructuring**, or the introduction of **indexes** or **materialized views** for **better performance**

AWS Data Engineer at WIPRO
Project 2: **08/2021-12/2022**
National Stock Exchange of India Limited

- Worked on implementing scalable infrastructure and platforms for **large-scale data ingestion**, **aggregation, integration, and analytics** in **Hadoop** using **Spark and Hive.**
- Architected a **cloud-based** data lake on **Amazon S3**, centralizing raw and processed data for seamless access to **analytics.**
- Designed and implemented **ETL workflows** using **AWS Glue** and Step Functions, reducing processing times by 45% and enabling efficient data migration.
- Built real-time data ingestion pipelines using **Apache Kafka** and **AWS Lambda**, supporting near real-time reporting with reduced **latency of 40%.**
- Integrated and transformed large datasets into **Amazon Redshift**, optimizing query performance and improving **analytics** capabilities by **35%**
  Orchestrated data workflows using **Airflow** and automated incremental loading into **Redshif**t for scalable and efficient reporting pipelines