# CHARAN LOKKU

Richardson, Texas
+1-469-558-0098 cxl220029@utdallas.edu LinkedIn GitHub

## EDUCATION

**The University of Texas at Dallas** *Richardson, Texas*
*Master of Science, Computer Science* **Jan 2023 - Dec 2024**

**SRM Institute of Science & Technology** *Chennai, Tamilnadu*
*Bachelor of Technology, Computer Science and Engineering* **Jul 2018 - May 2022**

## RELAVENT EXPERIENCE

**The University of Texas at Dallas** *Richarson, Texas*

**Stock Price Prediction Using ARIMA and PySpark** **Apr 2024 - May 2024**

- Processed large-scale stock data using Spark, handling missing values, and normalizing timestamps.
- Performed EDA and stationarity testing with Spark Data Frame operations and the ADF test for ARIMA suitability.
- Implemented ARIMA, optimizing p, d, q parameters and applying time-series decomposition for trend analysis.
- Trained and validated predictions using MAE and RMSE, ensuring forecasting accuracy.
- Visualized stock trends with Matplotlib and Seaborn and optimized performance using Spark's distributed computing.

**Real-Time News Data Pipeline Using Kafka, Spark Streaming, and ELK** **Apr 2024 - Apr 2024**

- Ingested real-time news data from NewsAPI, filtering articles by keywords, sources, language, and time, and streamed them into Kafka for real-time processing.
- Processed real-time data using PySpark and Spark Streaming, applying filtering, aggregation, and Named Entity Recognition for entity extraction.
- Indexed processed data in Elasticsearch, utilizing Logstash for structured ingestion and enabling efficient search and analytics.
- Designed Kibana dashboards to visualize top named entities and real-time trends, tracking mentions at 15, 30, 45, and 60-minute intervals.

**Analyzing Social Networks Using GraphX/Graph Frame** **Mar 2024 - Mar 2024**

- Conducted graph-based computations and analyses using GraphX/Graph Frame in PySpark, identifying key nodes, calculating PageRank, and finding largest components.
- Provided insights into social network structure, demonstrating proficiency in distributed computing and data analysis for decision-making.

**Web Crawling & NLP-Based Chatbot for Personalized Information Retrieval** **Feb 2024 - Mar 2024**

- Developed a web crawler using Beautiful Soup with URL filtering and error handling for structured data extraction.
- Processed text with NLTK and applied TF-IDF (sklearn) to build a structured knowledge base stored as a pickle file.
- Built a chatbot with NLP tokenization, rule-based query matching, and a user model for personalized responses.
- Implemented fallback mechanisms for unrecognized queries to enhance engagement.
- Utilized Python, Beautiful Soup, Requests, NLTK, Scikit-learn, Pandas, and Pickle for development.

**KPMG - *Data Analytics Consulting Internship Program (Forage)*** **Apr 2021 - Jun 2021**

- Conducted customer segmentation and predictive analysis to assess purchasing behavior, loyalty, and churn risks.
- Used SQL for data extraction and cleansing and Python for preprocessing, handling missing values, and ensuring data accuracy.
- Implemented RFM analysis and quantile-based ranking to classify customers for targeted marketing and retention.
- Built a churn prediction model using machine learning to identify at-risk customers.
- Developed Tableau and Power BI dashboards to visualize trends, industry-specific purchases, and regional spending, driving data-driven decisions.

## TECHNICAL SKILLS

**Programming Languages:** Python, SQL, R
**Tools & Technologies:** Apache Spark, Hadoop, Kafka, Docker, NumPy, Pandas, Databricks, AWS, MySQL, MongoDB, Tableau, Power BI, Git, Apache Airflow, Snowflake, Matplotlib, Seaborn, Version control, NoSQL databases
**Big Data Skills:** Real-time processing, ETL workflows, data warehousing

## COURSES AND CERTIFICATIONS

**Cognitive computing with IBM cloud**: A workshop offered by IBM & approved by SRM IST **Oct 2020**
**Machine Learning**: Course organized by Stanford University and offered through Coursera **May 2020**

## RESEARCH AND PUBLICATION

**Research Paper:** Classification of Genuinity in Job Posting using Machine Learning
**Journal Name:** *International Journal for Research in Applied Science & Engineering Technology* **(IJRASET)** **Dec -2021**