

KARTHIK RAJA

Boston, MA | raja.ka@northeastern.edu | [LINKEDIN](#) | [GITHUB](#) | [PORTFOLIO](#) | +1-(781)-921-6685

EDUCATION

Northeastern University, Boston, USA

Master of Science in Data Analytics

Expected May 2025

GPA: 3.8

Relevant Coursework: Database Management and Design, Statistics, Computation and Visualization

CMRIT, Bengaluru, India

August 2021

Bachelor of Engineering in Electronics & Communication Engineering

SKILLS

- **Programming Languages**: Python, Java, C, JavaScript, SAS, R, Power Shell
- **Visualization Tools**: Tableau, PowerBI, MS Excel, Flask, Git, Jira, MS PowerPoint, Matplotlib, Seaborn
- **ML/AI**: SVM, RCNN, LSTM, Pytorch, Sklearn, Scipy, NLTK, OpenCV, Tensorflow, Mediapipe
- **Cloud Services**: Databricks, AWS Elastic MapReduce, EC2, S3, Redis, Lambda, GCP, Azure Synapse Analytics, Data Studio
- **Databases**: SQL, PostgreSQL, MS SQL Server, Mongo DB, Hadoop, Spark Streaming, Kafka, Snowflake, Cassandra
- **Skills**: Data Pipelines, Statistical Modelling, Agile, Hypothesis A/B Testing, Rest API, Communication, Presentation, OOPS
- **Certificates**: Data Science with Python: IBM Certified(2027292), Machine Learning Advanced: IBM Certified(2136390), SAS Base 9.4, SQL, Cloud Computing

EXPERIENCE

Tata Consultancy Services, Chennai, India

August 2021 - August 2023

Data Engineer

- Built ETL pipelines to load terabytes of streaming data from various upstream modules to a data warehouse in real-time using Spark and **Apache Kafka** on **AWS EMR** clusters, resulting in a 20% increase in data processing speed
- Worked with REST APIs to fetch historical data from legacy systems to push to Kafka and capture changes in data updates
- Constructed NoSQL data aggregation pipelines in MongoDB and dealt with ad hoc requests for data availability
- Established Agile processes by working with the leadership team to better team collaboration and ship features 30% faster

Northern Suites, Bangalore, India

August 2020 – December 2020

Data Analyst Intern – Python, SQL, PowerBI

- Built out the data and reporting infrastructure from the ground up using SQL Server, Python, and PowerBI to provide real-time insights on guest data to non-technical stakeholders, increasing customer insights by 18%
- Designed and implemented A/B experiments to optimize marketing campaigns, raising click-through rates by 27%
- Employed data manipulation techniques in SQL to transform raw marketing data, which improved quality by 52%
- Deployed machine learning models to predict customer behavior, leading to a 19% increase in customer engagement

Folk Developers, Bangalore, India

December 2019 – May 2020

Machine Learning Intern - Python(NLP, Flask, Pandas, Numpy, Plotly), AWS RDS, QuickSight

- Restructured existing schema using **AWS RDS**, and identified **KPIs** using regression, increasing customer insights by 15%
- Designed **NLP-based Twitter sentiment** analysis bot using **Python**, achieving 70% market sentiment accuracy, Link-[GitHub](#)
- Utilized Plotly to craft **interactive customer insights dashboards**, obtaining a 10% surge in customer support feedback
- Collaborated cross-functionally to scrape and analyze data towards the development of the 'Bhagavad Gita' chatbot

PROJECTS

H1B Data Engineering, Personal Project - Python, Tableau Link - [GitHub](#)

September 2023 - October 2023

- Scraped and cleaned over 5M CSV data records(2009 – 2023) from USICS website using Python to analyse H1B acceptance
- Built a pipeline to extract data and transform it using Python API, orchestrated the extraction timing using Airflow
- Aggregated and analysed list of companies with highest number of H1B sponsorship, visualized the insights using Tableau
- Resulted in valuable insights matching with 4 different sources, leading to a 95% accurate H1B acceptance trend analysis

Yoga with ML/AI, Northeastern University - Python, MediaPipe, Link - [GitHub](#)

October 2023 - January 2024

- Innovated an application to detect yoga poses and provide estimation on accuracy of poses using Mediapipe library
- Collected and cleaned data over 1000 images of training and testing data for more than 7 yoga poses, trained the model to classify between different poses using SVM and optimized the accuracy further by deploying RCNN and LSTM by 20%
- Provided active text feedback on estimation of poses using OpenCV, eliminating the need of a physical instructor