

# Karthikeyan Rajagopal

karthikeyanr305@gmail.com • +1-716-907-9288 • LinkedIn • Github • Website

## Professional Summary

---

Data Scientist with 4+ years of experience spearheading cross-functional teams and driving growth for global clients such as Meta Inc., Emirates, HSBC, Mashreq, and Citibank. Expert in transforming complex business problems into data-driven strategies and actionable insights through machine learning, predictive modeling, analytics, and visualizations.

## Technical Skills

---

**Programming Language and Development:** Python, SQL, Java, Scala, R, C, C++, Flask, Git

**Data Viz and Analysis:** Power BI, Tableau, Looker, Matplotlib, Seaborn, Plotly, Streamlit, Exploratory Data Analysis (EDA), Business Analysis, Business Intelligence, Statistical Modeling

**Statistics:** Regression, Classification, Hypothesis Testing, Causal Inference, A/B Testing, Bayesian Inference

**Big Data and Cloud:** Pandas, Numpy, Scikit-learn, PyTorch, TensorFlow, PostgreSQL, MySQL, AWS, Azure, GCP, Spark, Hadoop, Apache Airflow, Redshift, BigQuery, Alteryx, DBT, Data Modeling

## Professional Experience

---

### Data Scientist — Crayon Data — Chennai, India

Jun 2018 - Jan 2021

#### Built client specific predictive models to increase sales and grow portfolio

- Enhanced recommendation engine using Python and SQL for feature engineering, A/B Testing and causal analysis for a fintech client, boosting user engagement by 23% across a 400+ merchant B2C two-sided marketplace
- Drove portfolio growth by \$1M for a global bank through an advanced user segmentation model through K-Means clustering algorithm, ideated personalized marketing strategies to increase Customer Lifetime Value(CLV)
- Deployed order demand forecasting using XGBoost and ARIMA in BigQuery for a cloud kitchen to maximize inventory utilization, DBSCAN clustering to optimize operations and transportation logistics, improving revenue by \$370K

#### Product Analytics to uplift Customer Engagement and Campaign Performance

- Led marketing campaigns using predictive analytics, designed Tableau dashboards to enhance market analysis, consumer insights, channel optimization, and campaign timing, yielding a 31% in ROI for a finance client
- Implemented over 20 A/B tests for an e-commerce SaaS marketplace to refine product features and requirements, achieving a 30% boost in Conversion Rate and a 15% lift in Average Order Value(AOV)
- Managed collaboration with 10+ cross-functional stakeholders to design product roadmaps and end-to-end user journeys, built Power BI Dashboards to monitor key product metrics, KPIs, and user behavior of 200,000 customers

#### Big Data Management and Advanced Analytics on large datasets

- Spearheaded a team of 5 to automate a Machine Learning ETL pipeline for deploying scalable models in AWS leveraging Scala, PySpark, and Hadoop; reducing time and resources in production by 54% and cost by 88%
- Scraped over 10,000 customer reviews from 50+ websites via Selenium, and preprocessed into a JSON based structured format for tag-based personalized recommendations, mitigating Churn Rate by 36%

### Data Science Research Assistant — SUNY Research Foundation — Buffalo, USA

Apr 2023 - Present

#### NLP Recommender System for Domain-Specific Recommendations in Research Conferences

- Extracted over 15,000 research papers via OpenAlex API, conducted data mining, cleaning, and wrangling, and engineered an optimized platform architecture with FAST API and MongoDB for scalable application development
- Leveraged Large Language Models (LLMs) such as SciBERT and developed an application based on semantic similarity and knowledge graphs using Transformers and Spacy, improving accuracy by 23% compared to SOTA
- Created a GenAI chatbot with a streamlined RAG (Retrieval-Augmented Generation) system and Llama-13b for content generation using Hugging Face, achieving a 0.85 BERT F1-Score and improved user experience

## Projects

---

### Credit Fraud Detection in Financial Transactions - Automated Prediction System [\[Link\]](#)

2024

- Pioneered a full-stack interactive platform leveraging Streamlit to predict fraudulent transactions with classification models such as Naive Bayes, Random Forest, XGBoost, Neural Network with SMOTE attaining an accuracy of 87%

### Stock Price Prediction using Time Series Forecasting and Deep Learning

2023

- Experimented with time-series forecasting techniques including ARIMA, Prophet, vanilla LSTM, and Multi-Stepped Stacked LSTM to accurately predict stock prices, yielding a best Root Mean Square Error (RMSE) of 4.98

## Education

---

Masters of Science in Computer Science & Engineering—University at Buffalo

2022 - 2024

Graduate Diploma in Data Science —University of Melbourne

2021 - 2021

Bachelor of Technology in Automotive Engineering—Indian Institute of Technology Madras

2013 - 2018