Karthikeyan Rajagopal

New York, United States (Open to Relocation) | +17169079288 | karthikeyanr305@gmail.com | LinkedIn | GitHub | Website

EDUCATION

University at Buffalo, **State University Of New York |** MS in Computer Science (GPA 3.81/4.00) **|**Buffalo, USA **AUG 2022 – DEC 2023**
University of Melbourne | Graduate Diploma in Data Science **|**Melbourne, AUSTRALIA **MAR 2021 – DEC 2021**
Indian Institute of Technology Madras (IIT M) |B. Tech & M. Tech (Automotive Engineering) **|**Chennai, INDIA **JUL 2013 – MAY 2018**

SKILLS & LEADERSHIP

Programming Language: Python, SQL, R, Java, Scala
Technology: Power BI, Tableau, Excel, Git, AWS, GCP, JIRA, Spark, Airflow, Hadoop, Redshift, Athena, BigQuery, Alteryx, Flask, SAS
Libraries: Pandas, Numpy, Matplotlib, Seaborn, scikit-learn(sklearn), TensorFlow, PySpark, PyTorch
Technical Skills: Natural Language Processing, Machine Learning, Exploratory Data Analysis (EDA), Data visualization, Statistics, Databases, Data Modelling, Regression models, Product Analytics, Causal Inference, Business Intelligence, Requirements Gathering

PROFESSIONAL EXPERIENCE

Research Assistant – Data Science | SUNY Research Foundation | Buffalo, New York **APR 2023 – PRESENT**

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

- Leveraged Large Language Models (LLMs) like SciBERT and optimized a recommendation platform for semantic similarity using RESTful APIs. Trained on CNN, BiLSTM, and Transformer models, improving accuracy by 23% compared to SOTA
- Integrated a RAG (Retrieval-Augmented Generation) system to enable chat interface using Llama-13b for content generation with FAISS as vectorDB, achieved a 0.85 BERT F1-Score and a 0.73 Spearman Coefficient against human scores using Llama 70b

Data Scientist | Crayon Data | Chennai, India **JUN 2018 – JAN 2021**

Generated insights to enhance Customer Engagement and Campaign Performance:

- Developed and managed Excel reports and interactive dashboards using Power BI and SQL to monitor business metrics, key performance indicators (KPIs) and user behavior of 200,000 customers enhancing customer insights and business processes
- Conducted pre-campaign statistical analysis using Python and SQL to identify right target audience. Developed Tableau dashboards extracting insights for ideal channels, timing, and frequency, boosting ROI by 31% for a finance client
- Spearheaded collaborative initiatives with 10+ cross-functional stakeholders across Finance, Marketing, Product and GTM teams and designed product roadmap to improve end-to-end user journey, facilitating communication and decision-making

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

- Utilized python for advanced feature engineering, A/B Testing and causal experimentation, boosting customer coverage by 300% for a global bank and improving user engagement (Click Through Rate - CTR) by 23% across a 400+ merchant marketplace
- Deployed order demand forecasting using XGBoost and ARIMA for a cloud kitchen in BigQuery to enhance inventory utilization, implemented DBSCAN clustering to optimize operational efficiency and transportation logistics, improving revenue by $370K
- Designed an advanced user segmentation model using K-Means clustering algorithm for a global bank, ideating targeted marketing strategies to increase Customer Lifetime Value (CLV), driving portfolio growth by $1M

Big Data Management and Advanced Analytics on large datasets:

- Led a team of 5 to architect and implement an automated ML ETL pipeline in AWS, leveraging technologies like Scala, PySpark, and Hadoop. Enhanced data throughput and efficiency, reducing time and resources in production by 54% and cost by 88%
- Scraped over 10,000 customer reviews from 20+ websites using Selenium, and preprocessed into a JSON based structured format for tag-based personalized recommendations, achieving a 37% increase in online user engagement.

ACADEMIC PROJECTS

Stock Price Prediction using Time Series Forecasting and Machine Learning | University at Buffalo **MAY 2023 – DEC 2023**

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

Credit Fraud Detection in Financial Transactions: Automated Prediction System | University at Buffalo **JAN 2023 – MAY 2023**

- Developed a full-stack interactive platform using Streamlit to predict fraudulent transactions using classification models such as Naive Bayes, Random Forest, XGBoost, Neural Network and handled class imbalance with SMOTE achieving an accuracy of 87%