Karthikeyan Rajagopal

New York, United States | +17169079288 | karthikeyanr305@gmail.com | LinkedIn | GitHub | Website

PROFESSIONAL SUMMARY

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

SKILLS & LEADERSHIP

Programming Language: Python, Java, SQL, Scala, C, C++, R, Go
Technology: Git, AWS, Azure, GCP, JIRA, Spark, Hadoop, Apache Airflow, Hive, Kafka, Power BI, Tableau, Flask, Docker, MongoDB
Libraries: Pandas, Numpy, Matplotlib, Seaborn, scikit-learn(sklearn), TensorFlow, PySpark, PyTorch, Spacy, Huggingface, LangChain
Technical Skills: Machine Learning, Natural Language Processing, Recommender Systems, Exploratory Data Analysis (EDA), Analytics, Data visualization, Statistics, Databases, Forecasting, A/B testing, Causal Inference, Dashboard, ETL, Geospatial Analysis

PROFESSIONAL EXPERIENCE

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

Built client specific applications to increase productivity and personalization:

- Enhanced proprietary recommendation engine via strategic feature engineering and experimentation, boosting customer coverage by 300% for a finance client and improving Click-through rate (CTR) by 23% across a 400+ merchant marketplace.
- Deployed order demand forecasting using XGBoost and ARIMA for a cloud kitchen; implemented DBSCAN clustering algorithm for delivery route optimization, yielding a $370K revenue increase and 36% better resource utilization.
- Designed an advanced K-Means clustering algorithm for a global bank through dynamic customer segmentation and targeted marketing strategies to increase Customer Lifetime Value (CLV), driving portfolio growth by $1M.

Refurbished an in-house end-to-end ETL and Machine Learning pipeline:

- Spearheaded a team of 5 to develop a Machine Learning big data pipeline for deploying scalable models on AWS with hyper-parameters to fine tune each module; reducing time and resources in production by 54% and cost by 88% per run.
- Streamlined data preprocessing with advanced entity resolution techniques using n-gram and fuzzy matching algorithms to achieve a reduction ratio of 3000:1, increasing accuracy and precision of predictions by 15%.

Improved Travel Recommendations for one of the Biggest Airlines in the world - Emirates:

- Collaborated with 10+ cross-functional stakeholders to improve end-to-end customer journey; designed interactive Power BI Dashboards to monitor key performance indicators (KPIs) of 20,000 customers.
- Leveraged sentiment analysis for tag extraction and scoring from historical travel data, integrating these into a matrix factorization-based collaborative filtering model, which boosted online user engagement by 37%.

Research Assistant – Machine Learning| SUNY Research Foundation | Buffalo, New York **APRIL 2023 – PRESENT**

Optimization of Microservices Architecture for Domain-Specific Recommendations:

- Utilized NLP techniques with Large Language Models (LLMs) to refine semantic similarity, resulting 32% surge in data extraction.
- Automated a Recommender System with RESTful APIs leveraging SciBERT to improve domain specific recommendations by 17%
- Leveraged FAISS as a dedicated vector database to house embeddings and execute similarity-based search operations.

ACADEMIC PROJECTS

Continuous Checkpointing for Deep Learning Recommender Systems | University at Buffalo **JAN 2023 – MAY 2023**

- Collaborated with Meta Inc. to develop a novel checkpoint mechanism, leveraging Kafka and PyTorch Hooks in a distributed AWS EC2 architecture for Terabyte dataset, which reduced recovery time from failures by 42% compared to SOTA.

Evaluating Multiple Machine Learning Classification Algorithms | University at Buffalo **JUL 2022 – DEC 2022**

- Enhanced classification accuracy in high-dimensional datasets by integrating PCA and LDA; benchmarked perceptron, decision tree, and neural network models using cross-validation; improved F1 score by 29% using L2 regularization and dropout.

Injury Severity Among Pedestrians: A Data Mining Approach | IIT Madras | Master's Thesis **AUG 2017 – MAY 2018**

- Analyzed pedestrian motor accident severity using statistical methods and logistic regression, coupled with feature engineering to identify key risk factors. Streamlined and improved accident risk predictions, achieving an AUC of 0.77 on the ROC curve.

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 2017 – MAY 2018**