ABHISHEK KUMAR

+44-7393068156 • abhishekkumar0598uk@gmail.com • www.linkedin.com/in/abhishek-kumar-a0190a18b • London, United Kingdom

Education

The University of Manchester — United Kingdom

Master of Science in Data Science - Distinction (Achieved) — 09/2023 - 09/2024

- **Relevant Coursework:** Statistics (Linear Algebra) and Machine Learning, Understanding Databases, Programming in Python for BA, Understanding Data & its Environment, Financial Data Analytics and AI in Finance, Text Mining.

SRM Institute of Science and Technology — India

Bachelor of Technology in Information Technology - Distinction (Achieved) — 05/2016 - 05/2020

- **Relevant Coursework:** Python, R Programming, Data Science and Big Data Analytics, DBMS, Data Mining, Data Warehousing, Statistics.

Experience

Data Scientist - Finance | MQube — 03/2024 - 09/2024

- Developed production-grade ML pipelines for financial market prediction using AWS SageMaker and Lambda, aligned with MLOps best practices.
- Fine-tuned domain-specific LLMs for sentiment analysis and returns forecasting, enhancing risk-adjusted returns by 30%.
- Implemented efficient retraining workflows using rolling window analysis on high-frequency time-series data.
- Effectively communicated model outcomes and tradeoffs to stakeholders including product managers, data engineers, and quant teams.

Data Analyst - Geospatial | Equans — 11/2023 - 02/2024

- Analyzed Scope 3 emissions for multiple travel modes (air, train, car, accommodation) using MySQL for data loading and EDA, ArcGIS, and Power BI, optimizing routes to reduce CO2 emissions by up to 75% and operational costs by 20%.
- Developed automated ETL pipelines and applied geospatial analysis (Shapely, GeoPandas, ArcGIS, NetworkX) to map high-emission routes, align corporate travel policies with sustainability goals, and drive data-driven decision-making for business decision-makers.

Data Scientist - Fraud and Risk | National Payments Corporation of India — 11/2022 - 08/2023

- Developed real-time and batch fraud detection systems using Random Forest and Autoencoder Deep Learning, reducing scam-related fraud by 35% and decreased false positive rates by 20% in a high-volume banking environment.
- Performed anomaly detection and fraud pattern analysis using Python and SQL, applying statistical methods to real-time financial transactions.
- Collaborated cross-functionally with product, compliance, and risk teams to align fraud strategies with business goals and minimize disruption.

Associate Software Developer | National Payments Corporation of India — 11/2020 - 11/2022

- Developed and optimized UPI application (Android & backend) using Java-based backend services using OOPs, Spring Boot, Android Studio, and integrated Kafka for real-time streaming and microservices communication, supporting 2 billion transactions monthly.
- Designed scalable payment solutions with A/B testing, ELK-based monitoring, and high-performance databases (Cassandra, Redis), while collaborating with cross-functional teams and external stakeholders to drive seamless integration and efficiency in an Agile environment.

Projects

Fraud Detection in Credit Card Transactions

- Built an unsupervised anomaly detection pipeline using Python, NumPy, Scikit-Learn, and unsupervised learning methods for fraud detection.
- Engineered robust feature scaling and PCA for dimensionality reduction, achieving 84% recall while minimizing false positives.

DocuChat-RAG

- Built an Agentic RAG-based chatbot using CrewAI, Qdrant (vector embeddings), LangChain and Streamlit for Q&A on PDFs & web search.
- Integrated local LLM inference (Ollama) and real-time fallback search (FireCrawl) for dynamic knowledge retrieval.

TextScout: Customer Transcript Analysis Using NLP and Generative AI

- Designed an NLP pipeline using RoBERTa, BART (HuggingFace), and FAISS for real-time sentiment analysis & zero-shot text classification.
- Implemented self-learning AI with automated retraining triggered by FAISS, improving model classification accuracy.

NLP-Driven Relation Extraction with Advanced Neural Networks

- Developed a deep learning model combining BiLSTM, CNN, and Multi-Head Attention to extract semantic relationships from text.
- Leveraged GloVe embeddings, NLTK for NLP preprocessing and TensorFlow/Keras to achieve a 75% F1 score on SemEval-2010 Task 8.

DiaRisk: Clinical Risk Assessment for Diabetes

- Performed in-depth exploratory data analysis and probability-based risk scoring to inform model design and interpretability.
- Owned the full ML pipeline: from EDA and model experimentation, designing PowerBI dashboard to MLOps integration using Docker and AWS.

Skills

Technical Skills:

Python (Pandas · NumPy · Matplotlib · Seaborn · PyTorch · OOP · scikit-learn) · TensorFlow · R · Java · PowerBI · DAX · Advanced SQL · PostgreSQL · MongoDB · DBT · Big Query · Machine Learning · Decision Trees · Random Forests · Time Series Forecasting · Predictive modeling · Anomaly Detection · Deep Learning · Neural Networks · Natural Language Processing (NLP) · LLMs · RAG · Prompt Engineering · AWS (Sagemaker · S3 · Lambda · Quicksight) · Microsoft Azure · Databricks · Version Control (Git) · Docker · MLOps · MLFlow · CI/CD · FastAPI.