Ajay John Alex, Data Scientist

07587335616, ajay.alex96@gmail.com

linkedin.com/in/ajay-john-alex96/

I am a results-driven Data Scientist with over 4 years of experience and a recent MSc graduate in Data Science from Nottingham Trent University, UK. I specialise in leveraging machine learning models to derive actionable insights across diverse industries. Experienced in optimising machine learning models and creating self-service analytics solutions to solve complex business problems. Proficient in Python, SQL, TensorFlow, PySpark, MLFlow, Azure services, and advanced pricing analysis. I attend stakeholder meetings independently, collaborate with teams to meet ad hoc data requests and maintain high standards of accuracy, written and verbal communication, attention to detail, timeliness, and business acumen. Strong communicator and problem solver, adept at fostering collaborative environments to contribute valuable data insights that support decision-making processes.

Expertise

- Python (PyTorch, NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn (sklearn)).
- Statistical Analysis (Hypothesis Testing, T-test, Chi-Square test, ANOVA).
- Data Analysis Tools (IBM SPSS, Microsoft Excel, KNIME).
- Data Visualisation, Dashboard design and Reporting (Tableau, Microsoft Power BI).
- Microsoft SQL Server Business Intelligence (BI) tool suite (SSIS, SSAS and SSRS, SAP B).
- Proficiency in cloud-based data warehouses including Azure Data Factory, and Snowflake environment.
- Data Mining (clustering and predictive modelling - regression and classification).
- Time Series Forecasting (ARIMA, SARIMA, Exponential Smoothing, LSTM networks).
- Database Management and Designing (RDBMS, MS SQL, T-SQL, DAX, ORACLE PL/SQL, Power Query, MySQL, BigQuery).
- Big Data Analytics (Kafka, Apache Spark, Databricks – PySpark, Spark SQL).
- Cloud Computing (Microsoft Azure, GCP).
- Proficiency in tracking and organising machine learning experiments, including metrics, parameters, and artifacts using MLFlow.
- Natural Language Processing NLP (Sentiment Analysis, Text Mining, Word Embedding).

Professional Experience

Data Analyst – Business School, Nottingham Trent University

October 2023 – Present

- Developed a predictive model using logistic regression to improve Higher Education outcomes. By analysing student demographics and academic data, I computed probability scores for achieving a 2:1 grade or higher, facilitating better retention strategies and academic support interventions.
- Utilize advanced data visualisation tools (e.g., Tableau, Power BI) to design intuitive and interactive dashboards for monitoring key performance indicators (KPIs) related to student enrolment, retention, graduation rates, financial performance, and institutional research.
- Demonstrated expertise in leveraging SQL Server Integration Services (SSIS) for highly efficient Extract, Transform, and Load (ETL) processes, leading to the generation of accurate and insightful reports.

Data Engineer Consultant – Clicflyer

March 2022 – September 2023

- Developed and implemented an image segmentation model with YOLO for auto cropping, reducing manual cropping time by 90%.
- Increased Text Auto Fill Optimizer efficiency and accuracy by 20-25% and 15-20% respectively, using Azure search service with microservices architecture.
- Designed Image Comparison Tool, leveraging Azure functions, Queue and OpenCV, enhancing efficiency by 66% with 99.7% accuracy.
- Designed and implemented a real-time data processing pipeline using PySpark, Databricks, and Apache Airflow to handle large volumes of streaming data for a multinational retail company. The project aimed to provide timely insights into customer behaviour, optimise inventory management, and enhance decision-making processes.
- Implemented advanced pricing analysis methodologies, integrating market research and analytics to create dynamic pricing models. Initiated strategic pricing adjustments, enhancing competitiveness across multiple product lines.
- Implemented end-to-end data pipelines to centralise cloud data storage like Snowflake and enabled

analysis for optimising product distribution, forecasting demand, and improving operational efficiency.

Data Analyst – Clicflyer

March 2020 – Feb 2022

- Developed recommendation tool optimising offers based on frequency, price range, and product placement, enhancing retailers' planning capabilities for data-driven decision-making, impacting 250+ MENA retailers.
- Engineered text classification model to reduce data errors in SKU-to-brand mapping, achieving 97.5% brand and 97% product prediction accuracy, enhancing data integrity.
- Scoped and managed the "Flyer Project" - a historic offer collection system, expediting data delivery for 12 MENA countries within 2.5 months, reducing cost by 70%.
- Boosted outlier detection accuracy to 15-17% with metrics, logical checks and statistical methods in a new PowerBI report.
- Launched API for the "Buy Online" feature, driving a 15-18% user engagement increase.
- Conducted web scraping for 15+ FMCG KSA retailers, ensuring accurate "In stock" tagging for 'Buy Online' improving online purchasing.
- Utilised formal problem-solving techniques such as the Ishikawa diagram, SWOT analysis, and decision trees to enhance decision-making processes.

Education and Certifications

Data Scientist Certificate	DataCamp	2024
MSc. Data Science	Nottingham Trent University	2022 – 2024
B. Tech Computer Science	Maharshi Dayanand University	2015 - 2019

Key Achievements

- In my published paper **- Enhancing Pollinator Conservation towards Agriculture 4.0: Monitoring of Bees through Object Recognition**, I closely collaborated with Nottingham Trent University & Aston University to develop a trained model that is then packaged within an explainable AI interface to facilitate use by non-technical users such as expert stakeholders from the apiculture industry towards informing responsible consumption and production.
- Designed and implemented an end-to-end data pipeline automation system for a financial institution using Microsoft Azure. Utilized Azure Data Factory to orchestrate workflows, integrating data from various sources such as SQL Server, Azure Blob Storage, and third-party APIs. Implemented data transformations and cleansing using Azure Databricks (PySpark), ensuring data quality and consistency.
- Implemented a scalable big data analytics solution for customer segmentation and targeting using Apache Spark and BigQuery. Processed and analysed customer data from multiple channels, including web traffic, social media interactions, and transaction history. Used unsupervised learning algorithms such as K-means clustering to identify distinct customer segments.
- Employed advanced natural language processing, including sentiment analysis and text mining, to develop a model that flags online Hate Speech. This in-depth study uncovered sentiments and recurring themes enriching our comprehension of language dynamics and their influence on public discourse.

References

Academic and professional references available on request