

ALLAN CHEERAKUNNIL ALEX

Data Scientist | Mumbai, India | +91 9930703439 | allanalex143@gmail.com
[LINKEDIN](#) | [GITHUB](#) | [PORTFOLIO](#) | [WEBSITE](#)

PROFESSIONAL SUMMARY

Results-oriented Data Scientist with a Master's in Computer Science and 2+ years of exposure to **Python, SQL, and Predictive Modeling**. Expert in engineering end-to-end ML pipelines (ETL, EDA) that drive business efficiency. Skilled in translating raw datasets into actionable insights using statistical analysis and visualization, reducing decision-making time by ~20%. Proficient in **Generative AI, FastAPI, and CI/CD** with a proven track record of deploying scalable ML applications.

TECHNICAL SKILLS

GenAI & ML: Generative AI, RAG, LangChain, LLMs, NLP, XGBoost, Deep Learning
Engineering: FastAPI, Streamlit, Render, CI/CD, Git, AWS (EC2, S3)
Libraries: Pandas, NumPy, Scikit-Learn, TensorFlow, FAISS, SHAP, TF-IDF
Languages: Python, SQL, JavaScript, HTML5, CSS3

PROFESSIONAL EXPERIENCE

Full Stack Data Science (Apprenticeship) - AlmaBetter Jun 2025 – Present

- Engineered scalable Machine Learning models for 3+ real-world production use cases, enhancing predictive capabilities.
- Orchestrated end-to-end ETL pipelines handling 10,000+ data points, including cleaning, feature selection, and tuning.
- Deployed multiple ML applications using FastAPI and Streamlit via CI/CD pipelines, ensuring 99.9% availability.

Data Science Intern - The Sparks Foundation Mar 2024 – Jun 2024

- Automated data quality checks using Python scripts, improving dataset accuracy by 25% and reducing manual errors.
- Formulated regression and classification models, achieving optimal performance metrics (RMSE, F1-score) on 5+ datasets.

Data Analyst Intern - Mentorless Mar 2024 – Jun 2024

- Designed interactive Power BI dashboards to track KPIs, visualizing 50+ key metrics for stakeholder review.
- Executed complex SQL queries to automate reporting, cutting manual data retrieval time by 30%.

Independent Research & Upskilling - Self-Directed Jan 2022 – Feb 2024

- Sharpened analytical reasoning and logic through 2000+ hours of UPSC preparation and problem-solving exercises.
- Mastered foundational concepts in Python, SQL, and Statistics through 10+ completed mini-projects.

Associate (Operations) - British Airways (WNS) Dec 2019 – Dec 2021

- Streamlined workforce reporting systems, improving operational efficiency by ~15% across the department.
- Directed data-driven analysis to minimize reporting delays, saving approximately 5 hours per week.

KEY PROJECTS

RAG-Based Document Intelligence Bot (GenAI, LangChain) Nov 2025 – Dec 2025

- Built a live GenAI application allowing users to chat with PDF documents using **Google Gemini 2.0 Flash**.
- Deployed on Streamlit Cloud with a RAG pipeline utilizing **FAISS** vector retrieval for context-aware answers.
- [Live App](#) | [View Code on GitHub](#)

AI-Powered Course Recommender (FastAPI, Streamlit) Sep 2025 – Oct 2025

- Architected a decoupled recommendation system using **FastAPI** (backend) and **Streamlit** (frontend).
- Engineered content-based filtering logic using **TF-IDF** and **Cosine Similarity** on 3,000+ courses.
- Implemented CI/CD pipelines via GitHub to deploy the API on **Render** and frontend on Streamlit Cloud.
- [Live App](#) | [View Code on GitHub](#)

Customer Churn Prediction (XGBoost, Explainable AI) Jul 2025 – Aug 2025

- Developed a churn prediction pipeline using **XGBoost** on 7,000+ records to identify attrition drivers.
- Optimized decision thresholds to maximize **Recall (55%)**, capturing high-risk customers for retention.
- Implemented **SHAP** (Explainable AI) to reveal contract type as the primary risk factor.
- [View Code on GitHub](#)

EDUCATION

M.Sc. Computer Science
Mumbai University | 2021
B.Sc. Computer Science
K.M. Agrawal College | 2018

CERTIFICATIONS

- Stanford Online: Intro to Statistics
- IBM: What is Data Science?
- CISCO: Data Analytics Essentials
- SAP: Generative AI