ALLAN CHEERAKUNNIL ALEX

Data Scientist │ Mumbai, India │ +91 9930703439 │ allanalex143@gmail.com

LINKEDIN │ GITHUB │ PORTFOLIO │ WEBSITE

PROFESSIONAL SUMMARY

Results-oriented Data Scientist with 2+ years of expertise in **Analytical Translation** and Python-based Machine Learning. Skilled in converting complex business contexts into scalable **Regression and Classification** models. Expert in engineering robust ML pipelines (ETL, EDA) and performing **Model Optimization** to drive decision-making. Proficient in **Generative AI, FastAPI, and Git** with a proven track record of solving real-world challenges through quantitative analysis.

TECHNICAL SKILLS

GenAI & ML:	**Regression**, **Classification**, **Optimization**, Generative AI, RAG, XGBoost, NLP
Engineering:	**Python (Scalable Code)**, FastAPI, Streamlit, Render, **Git Versioning**, CI/CD, AWS
Libraries:	Pandas, NumPy, Scikit-Learn, TensorFlow, FAISS, SHAP, TF-IDF
Tools/Langs:	SQL, Excel, PowerPoint, JavaScript, HTML5, CSS3

PROFESSIONAL EXPERIENCE

Data Science Intern - *The Sparks Foundation*   Mar 2024 – Jun 2024
- **Automated** data manipulation tasks using Python, 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 - *Mentorness*   Mar 2024 – Jun 2024
- **Designed** interactive Power BI dashboards to track KPIs, utilizing Excel and SQL for quantitative analysis.
- **Collaborated** with stakeholders to unravel data challenges, cutting manual data retrieval time by 30%.

Independent Research & Upskilling - *Self-Directed*   Jan 2022 – Jun 2023
- **Sharpened** quantitative mastery and logic through 2000+ hours of problem-solving exercises.
- **Mastered** foundational concepts in Python, SQL, and Statistical Methods through 10+ completed mini-projects.

Associate (Operations) - *British Airways (WNS)*   Dec 2019 – Dec 2021
- **Streamlined** workforce reporting systems, fostering relationships with diverse stakeholders to meet deadlines.
- **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 for analytical translation, allowing users to chat with PDF documents using **Google Gemini**.
- **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** and **Git versioning** for modular code structure.
- **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, Model Optimization)   Jul 2025 – Aug 2025
- **Developed** an end-to-end classification pipeline using **XGBoost** to solve the business challenge of customer attrition.
- **Optimized** decision thresholds to maximize **Recall (55%)**, capturing high-risk customers for retention.
- Implemented **SHAP** to interpret model decisions, ensuring solutions are anchored in statistical methodologies.
- View Code on GitHub

EDUCATION

Master of Science in Computer Science (Specialization: AI & ML) │ Woolf University (via AlmaBetter)	Jun 2023 – Jun 2026 (Expected)
M.Sc. Computer Science (Specialization: Data Science) │ Mumbai University	2021
B.Sc. Computer Science │ K.M. Agrawal College	2018

CERTIFICATIONS

AlmaBetter: Full Stack Data Science Job-Ready Certificate │ **Stanford Online:** Intro to Statistics │ **IBM:** What is Data Science? │ **CISCO:** Data Analytics Essentials │ **SAP:** Generative AI