

KAAN ALTAY

Data Scientist | Analytics & Forecasting | ML, GenAI & LLM Systems

Istanbul, Turkey | kaan_altay@outlook.com.tr | [linkedin.com/in/kaan-altay](https://www.linkedin.com/in/kaan-altay) | <https://github.com/kaanaltay-hub>
www.kaanaltay.com |

PROFIL

Data Scientist and Machine Learning Engineer with a Statistics degree, financial auditing experience at PwC Turkey, and a track record of building production ML systems at ImeBrands (Illinois, USA). That combination — analytical rigour from the Big Four and hands-on ML engineering in e-commerce — gives me a perspective most data scientists don't have. Delivered end-to-end solutions across time series forecasting, NLP, Computer Vision, and recommendation systems using FastAPI, Docker, and AWS. I don't just build models; I build systems that solve real business problems.

TECHNICAL SKILLS

Programming & Frameworks: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Keras, TensorFlow, PyTorch, FastAPI, Flask, Streamlit), SQL (PostgreSQL, MySQL, MSSQL, Oracle SQL), Git & GitHub

Machine Learning & Deep Learning: Supervised Learning (Regression, Classification, Tree Models, Gradient Boosting, CatBoost, XGBoost), Unsupervised Learning (Clustering, PCA), Deep Learning (ANN, CNN, RNN/LSTM/GRU), Time Series Forecasting, Predictive Modeling, Feature Engineering

NLP & Generative AI: Natural Language Processing (NLP), Sentiment Analysis, LLMs, RAG Architecture, Generative AI, Prompt Engineering, AI Agents

Computer Vision & Recommendation Systems: CNN-based Image Classification, Computer Vision, Recommendation Systems, Collaborative Filtering

Statistics & Analytics: A/B Testing, Effect Size Analysis, Statistics I & II, RFM Models, CLV Estimation, Cohort Analysis, Churn Analysis, Exploratory Data Analysis (EDA)

Visualization & Business Intelligence: Power BI (DAX, Power Query), Tableau, Google Sheets, Excel, Data Storytelling

MLOps & Cloud: AWS (SageMaker, S3, RDS, Deployment), Docker, FastAPI (REST API), Streamlit, MLOps Pipelines, Jira, SDLC, SMOTE, Model Deployment

PROFESSIONAL EXPERIENCE

Data Scientist | **ImeBrands** — Illinois, USA (Remote)

Jan. 2025 – Feb. 2026

Time Series Forecasting & Business Optimization

- Developed complex time series models (ML and Deep Learning with TensorFlow/PyTorch) to forecast revenue and profitability — directly integrated into operational planning and strategic decision-making.
- Built and deployed a predictive ML model to forecast **Amazon Buy Box %** — improved product positioning and win rates; deployed production-ready model on AWS infrastructure.
- Implemented **inventory and demand optimization models** based on historical sales data and seasonal patterns — demonstrably reduced overstock costs.
- Optimized **Amazon PPC campaigns** through data-driven budget allocation and ROAS analysis — maximized ROI on advertising spend.

Machine Learning & Customer Analytics

- Designed and evaluated rigorous **A/B tests** and **effect size analyses** with statistically significant samples — evidence-based optimization of marketing strategies and conversion rates.
- Implemented **sentiment analysis** on thousands of product reviews using **NLP** — generated automated quality signals and product improvement insights.

AI Systems, Recommendation Systems & Computer Vision

- Developed a personalized **recommendation system** using collaborative filtering and purchase histories – unlocked cross-selling potential and increased basket value.
- Built an intelligent **chatbot with RAG architecture** (Retrieval-Augmented Generation) based on a SQL company database; developed agent-based AI assistants for automated market trend analysis.
- Led a **computer vision** project for CNN-based automatic classification of product categories from image data – replaced manual categorization processes with an AI solution.

Data Analysis & Business Intelligence

- Conducted multidimensional **revenue, profitability, and churn analyses** using SQL and Python – **root cause analysis** of canceled orders revealed logistics and product issues.
- Created interactive **KPI dashboards** in **Power BI** and **Tableau** for C-level reporting – transformed complex datasets into clear strategic decision-making tools.

Career Break

Aug 2023 – Dec 2024

Stepped away from work to focus on health. Used the period to make a deliberate transition into data science — self-directed learning, hands-on project development, and building the foundation that led directly to the Data Scientist role at ImeBrands.

Assurance Associate | PwC Turkey — Eskisehir, Turkey

Sep. 2022 – July. 2023

- Analyzed large-scale financial datasets including trial balances and sub-ledgers to support fund audit and IFRS reporting processes.
- Performed data validation and reconciliation checks to identify inconsistencies across financial reports for manufacturing and service clients.
- Reviewed and controlled year-end financial reports, including rate and variance calculations, ensuring data accuracy and compliance.
- Tracked and documented reconciliation processes to improve data transparency and audit traceability.
- Collaborated with cross-functional teams to resolve data quality issues and support timely reporting.
- Conducted year-end inventory and cash counts, verifying data completeness and consistency with reported figures.

SELECTED PROJECTS

Amazon Analytics Chatbot | Python · LangChain · RAG · SQL · Streamlit · HuggingFace

- Built a hybrid AI assistant that automatically routes questions between SQL query generation and semantic search across 6 integrated data sources covering revenue, advertising, and product metrics.
- Designed the RAG pipeline using LangChain and a local LLM on HuggingFace; exposed underlying SQL queries for full transparency and auditability.
- Result: Natural language interface that eliminates the need for SQL knowledge while maintaining full data transparency.

GitHub: <https://github.com/kaanaltay-hub> · Live Demo: [huggingface](https://huggingface.com)

Amazon Buy Box Prediction | Python · XGBoost · LightGBM · Scikit-learn · Streamlit

- Engineered 78 features including lag variables, rolling statistics, B2B ratios, and advertising data; trained and compared 6 models (Ridge, Random Forest, Gradient Boosting, XGBoost, LightGBM, Ensemble).
- Generated per-ASIN forecasts with confidence intervals and automatic risk flags (HIGH / MEDIUM / LOW).
- Result: Advertising presence emerged as the #1 predictor with 81% feature importance — directly shaped ad spend strategy.

GitHub: <https://github.com/kaanaltay-hub> · Live Demo: [huggingface](https://huggingface.com)

Sentiment Analysis — Product Reviews | Python · BERT · HuggingFace · KeyBERT · PyABSA · Streamlit

- Ran 4 NLP models in parallel: BERT for sentiment classification, a custom model for root cause categorization (shipping, price, quality), KeyBERT for key phrase extraction, and PyABSA for aspect-based sentiment analysis.
- Result: Transformed raw customer reviews into actionable product improvement signals — not just a sentiment score, but a diagnosis.

GitHub: <https://github.com/kaanaltay-hub> · Live Demo: [huggingface](https://huggingface.com)

Amazon Revenue Forecasting | Python · SARIMA · Prophet · XGBoost · LSTM · Streamlit

- Built a 4-model forecasting system — SARIMA for seasonality, Prophet for holiday effects, XGBoost with 78 engineered features, and LSTM for long-term dependencies — combined into a final ensemble.
- Generated weekly revenue and profit forecasts up to 13 weeks ahead, evaluated across MAPE, MAE, and RMSE.
- Result: End-to-end forecasting pipeline designed for quarterly business planning.

GitHub: <https://github.com/kaanaltay-hub> · Live Demo: [huggingface](https://huggingface.com)

Product Image Classifier | Python · TensorFlow · MobileNetV2 · Transfer Learning · Streamlit

- Fine-tuned MobileNetV2 on a frozen ImageNet backbone with custom top layers, data augmentation, and EarlyStopping across 9 product categories and 3,632 images.
- Achieved 82% validation accuracy with top categories exceeding 90%; supports both single-image and batch prediction modes.
- Result: Replaced manual product categorization with an automated Computer Vision pipeline.

GitHub: <https://github.com/kaanaltay-hub> · Live Demo: huggingface.co

Additional projects (Cross-Sell Recommendation System, Olist E-Commerce Analytics, Stroke Risk Analytics, Amazon Revenue & Profitability Analysis, Amazon Sales Analytics, Amazon Marketplace Analytics) available at: <https://kaanaltay.com>

EDUCATION

Bachelor of Science – Statistics

2016 – 2022

Eskisehir Technical University, Faculty of Science – Eskisehir, Turkey

- Relevant Coursework: Probability and Statistics, Statistical Inference, Data Analysis, Regression Analysis, Linear Algebra

Erasmus+ Exchange Program

2020 – 2021

Czestochowa University of Technology, Faculty of Science – Czestochowa, Poland

LANGUAGES

Turkish: Native

English: Intermediate