

BENEDICT BETT

Eldoret, Kenya (Remote) | 0718 647 283 | benedictbett08@gmail.com | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Scientist with proven experience building and deploying production-grade machine learning and predictive analytics systems. Developed a live churn prediction API achieving 84.2% ROC-AUC and a credit risk scoring platform trained on 307,000+ records — both using Python, scikit-learn, XGBoost, and LightGBM. Familiar with NLP fundamentals, transfer learning concepts, and LLM APIs. Proficient in the full ML pipeline from data processing and feature engineering to model deployment in production environments. Passionate about using AI to solve real-world problems at scale

TECHNICAL SKILLS

Programming & Analysis: Python (Pandas, NumPy, SciPy, scikit-learn), SQL (Advanced), R

Machine Learning & Predictive Analytics: XGBoost, LightGBM, Random Forest, Logistic Regression, Gradient Boosting, Feature Engineering, SHAP, SMOTE, Cross-Validation, Hyperparameter Tuning

NLP & Embeddings: NLP fundamentals, SentenceBERT (familiar), Gensim (familiar), transfer learning concepts, LLM APIs, text classification

Data Engineering: Large dataset processing, data pipelines, ETL workflows, automated cleaning

Deployment & Cloud: FastAPI, Streamlit, Docker, Render.com, REST APIs, GCP/BigQuery (familiar), AWS (basic)

Tools: Git, GitHub, Jupyter Notebook, VS Code, Matplotlib, Seaborn, Plotly

PROJECTS

Safaricom Churn Intelligence — Predictive Analytics & Production API | [Live API](#) | [GitHub](#)

- Built and deployed a production-grade predictive model achieving 84.2% ROC-AUC — using gradient boosting and ensemble techniques to forecast customer churn across 7,043 records.
- Engineered 25+ behavioral features from raw transaction data — demonstrating ability to extract meaningful signals from large, complex datasets.
- Deployed as a production FastAPI REST API with 6 endpoints, batch processing, and full OpenAPI documentation — pushing ML code directly into production.
- Projected 567% ROI and KES 800M+ annual savings — demonstrating ability to connect ML model outputs to real business impact.
- Designed end-to-end model architecture, feature selection, and deployment pipeline in a self-directed environment.

CreditIntelligence — Large-Scale Predictive Scoring System | [Live App](#) | [GitHub](#)

- Built an end-to-end predictive scoring system trained on 307,000+ records (ROC-AUC: 0.75) – processing large datasets using Pandas, NumPy, and SciPy.
- Applied XGBoost, LightGBM, Random Forest, and Logistic Regression with cross-validation and hyperparameter tuning – selecting optimal model for production deployment.
- Implemented SHAP explainability to surface key prediction drivers and support transparent, interpretable decision-making.
- Deployed as a live Streamlit application with real-time scoring and automated recommendations.

EXPERIENCE

Data Science & Analytics Intern – Future Interns Fellowship Program March 2026 – Present | Remote

- Building and deploying ML models and predictive pipelines on real-world datasets.
- Applying Python, NumPy, Pandas, and scikit-learn to develop production-ready analytics solutions.
- Collaborating with a global cross-functional remote team across multiple time zones.

EDUCATION

Diploma in Computer Science – Eldoret Polytechnic 2024 – 2026 | Data Structures, Algorithms, Statistics, Database Systems, Software Development

Applied Data Science Lab – WorldQuant University 2026 – Present | In Progress

CERTIFICATIONS

- Machine Learning with Python – Coursera
- SQL (Advanced) – Kaggle
- Databases and SQL for Data Science with Python – Coursera (In Progress)