Mark John Patel

patelmarkjohn@gmail.com | 0746236079
LinkedIn | GitHub| Portfolio Website

Skills

- **Machine Learning & Model Deployment:**
  Skilled in machine learning algorithms, especially tree-based methods and logistic regression, with experience deploying predictive models at scale.
- **Python & SQL Proficiency:**
  Strong Python skills (pandas, NumPy, scikit-learn) for data analytics and modeling, paired with SQL expertise for efficient data management and analysis.
- **Experimental Design & Statistical Analysis:**
  Experienced in experimental design, statistical tests, and interpreting diverse data types for rigorous analysis.
- **Insightful Communication:**
  Ability to translate complex data insights into accessible information for stakeholders of all levels.
- **Attention to Detail & Team Collaboration:**
  Methodical, detail-oriented, and highly collaborative, with a focus on accuracy and teamwork.
- **Cloud & Parallel Processing Knowledge:**
  Familiar with cloud platforms (GCP, AWS) and tools like Spark and Dask for handling large datasets effectively.
- **Financial Data Experience:**
  Background in working with financial data and credit scoring models.

EXPERIENCE

Data Scientist
Applied Data Science Lab, Worldquant University *June 2024 – Present*

- Collaborated with a multidisciplinary team to implement real-world data science solutions, focusing on business and societal problems using machine learning and predictive analytics.
- Developed a **predictive model** for a financial institution, using Random Forest and Logistic Regression, which improved loan default prediction accuracy by 18%, enhancing risk management efforts.
- Led the development of an **automated data pipeline** using Python and PostgreSQL to process large datasets efficiently, reducing manual effort by 30%.
- Analysed complex datasets, creating **Tableau dashboards** and reports that provided actionable insights to stakeholders in the healthcare industry.
- Contributed to a project that used **time series forecasting** models to predict patient admission rates, assisting hospitals in optimizing resource allocation.

PROJECTS

Predictive Pricing Model for Kenyan Fuel Market
GitHub Link
Developed a machine learning model in Python using scikit-learn to predict fuel prices with 92% accuracy. Implemented a Random Forest algorithm and factored in inflation and exchange rates, resulting in a 5% cost-saving prediction accuracy.

April 2024 – May 2024

Credit Risk Assessment Using Machine Learning
GitHub Link
Created a credit risk model using Logistic Regression, Random Forest, and XGBoost. Improved accuracy by 15% with hyperparameter tuning using GridSearchCV, enhancing credit analyst efficiency by 30%.

June 2024 – July 2024

Breast Cancer Diagnosis Prediction Model
GitHub Link
This project implements a machine learning model using PyTorch to predict breast cancer diagnoses based on various medical features. The model uses a neural network classifier to distinguish between malignant and benign breast cancer diagnoses.

Dec 2024

EDUCATION

Bachelor of Science in Computer Science
South Eastern Kenya University *Aug 2019 – Oct 2023*

AWARDS

- APHRC DSE Inspire Hackathon 2024 Finalist *Feb 2024*

AFFILIATIONS

Applied Data Science Lab, Worldquant University *June 2024 – Present*