

# Theophilus Utobo

Data Scientist

[Obinnapt@gmail.com](mailto:Obinnapt@gmail.com)

(+44) 7405822120

[Website/Portfolio](#)

[LinkedIn](#)

United Kingdom

## PROFESSIONAL SUMMARY

An analytical and detail-oriented Data Scientist with extensive experience in Python, R, and SQL. Skilled in data mining, cleaning, modeling, and analysis to extract actionable insights that support data-driven decision-making. Recently enhancing my expertise through the **Data Science with Machine Learning & AI Career Accelerator** by the **University of Cambridge**, further expanding my knowledge in advanced machine learning and artificial intelligence concepts. Motivated about leveraging data to solve complex problems and drive innovation.

## PROJECTS

### [Predicting Student Dropout Using Machine Learning \(Study Group – Feb/2024\)](#)

- Developed a machine learning model to predict student dropout, achieving 99% accuracy and AUC of 1.00 using XGBoost and Neural Networks.
- Implemented a three-stage data analysis approach, identifying academic performance as the strongest predictor of student retention.
- Provided actionable recommendations, including real-time attendance tracking and automated alert systems, to improve student retention strategies.

### [Customer Segmentation Using Clustering Techniques \(University of Cambridge - \[Dec/2024\]\)](#)

- Conducted a detailed customer segmentation analysis using a dataset with over 950,000 records to identify actionable customer groups for targeted marketing.
- Applied K-Means clustering to group customers into five distinct segments, optimizing clusters using the Elbow Method and Silhouette Score.
- Visualized cluster separations and relationships using PCA and t-SNE, ensuring clear and interpretable insights.
- Delivered creative actionable strategies, including prioritizing high-CLV customers, targeting younger demographics, and re-engaging low-recency customers.

### [Anomaly Detection in Ship Engine Dataset \(University of Cambridge - \[Nov/2024\]\)](#)

- Conducted anomaly detection on a dataset of over 19,000 records to identify early indicators of ship engine malfunctions, focusing on six critical features such as Engine RPM and Lubrication Oil Temperature.
- Implemented and compared IQR, One-Class SVM, and Isolation Forest methods, achieving a 2.5% anomaly detection rate with Isolation Forest as the most effective approach.
- Visualized anomalies using PCA-based 2D plots to enhance interpretability and support stakeholder decision-making.
- Delivered actionable recommendations, prioritizing key metrics like Lubrication Oil Temperature for monitoring and targeted maintenance strategies.

## WORK EXPERIENCE

### **Particle Physics Department | IT Helpdesk (Oct 2018 – Sept 2019)**

- Managed and optimized **SQL databases**, ensuring efficient research data storage, retrieval, and security.
- Automated **data reporting processes** using Python and SQL, reducing manual work by 30%.
- Used Python and Pandas to **process log files** and extract key insights for network performance improvements.

## SKILLS

### **Programming Skills**

- Python (NumPy, Pandas, Scikit-learn, TensorFlow, Keras), R, SQL

### **Data Visualization**

- Matplotlib, Seaborn

### **Machine/Deep Learning**

- Supervised & Unsupervised learning, Regression, Classification

### **Tools & Technologies**

- Jupyter Notebook, Git, Excel

## EDUCATION

### **Data Science Career Accelerator**

University of Cambridge

September 2024 - Current

### **M.Sc.**

### **Data Science**

Nottingham Trent

University

September 2020 - June 2022

Nottingham

Distinction

### **B.Sc.**

### **Computer Systems**

Nottingham Trent

University

September 2016 - June 2020

Nottingham

1<sup>st</sup> Class Honors.

## RELEVANT COURSES

- Statistical Data Visualization and statistical analysis.
- Applied AI, Data Mining Data Processing.
- Software Engineering.
- Research Project.