

# THEURI GICHUKI

## DATA ENGINEER

### CONTACT



<https://g-theuri.github.io/Portfolio/>



<https://www.linkedin.com/in/theurigichuki/>



Nairobi, Kenya

### SKILLS

#### // PROFESSIONAL

ETL and Data Pipelines

Machine learning with Apache Spark

Data Warehousing and BI analytics

Big Data with Apache Spark and Hadoop

#### // TECHNICAL

Python

SQL & NoSQL

Shell Scripting

HTML & CSS

### EDUCATION

#### Certificate

IBM Data engineering  
Professional Certificate  
Coursera  
2024

[Verify this certificate](#)

#### Bachelor of science

Computer Technology  
Multimedia University, Kenya  
2021

### SUMMARY

I am a self-motivated data engineer specializing in ETL processes, data mining and big data processing with a strong background in computer science.

I have substantial knowledge in big data technologies such as Hadoop and Apache Kafka and a strong understanding of distributed systems and data processing frameworks software like Apache spark. I have good understanding in NoSQL databases like Cassandra and proficient with RDBMS technologies such as MySQL and PostgreSQL. I have significant knowledge in building data pipelines using Python, Bash Scripting, Apache Airflow and IBM Cloud, implementing Airflow DAGs to automate ETL processes.

Some of my recent projects involve web crawling/web scraping a publicly available e-commerce book webstore scraping 450+ records on a single category while considering pagination, then analyzing the scraped data. I have also worked on machine learning models leveraging Naïve Bayes classifier algorithms for the Titanic survival prediction with an accuracy score of 80%. These projects are available on my [portfolio](#).

I am proactive and a fast learner, constantly researching and evaluating progressive technologies for production and utilization. I have excellent communication and problem-solving skills and the ability to work under minimum supervision.

I am eager to leverage my expertise to provide data management solutions and strategic insights to drive business growth and efficiency.

### EXPERIENCE

#### Data Annotator

Nov 2020 – Present

CloudFactory, Nairobi

- Annotate and label image, video and audio using various internal software programs for machine learning purposes.
- Attaching metadata to image, video, audio and text data in accordance with the established procedures.
- Participate in addressing performance and development issues for automation process.
- Liaising with the management to document business processes accurately.

#### Computer technician

Jan 2020 – April 2020

Kenya School of Agriculture, Nyeri

- Developed a library management system software with 100+ users to facilitate library data maintenance.
- Developed an inventory software to keep records of the institution's computer hardware and gadgets.
- Standardized and optimized database models to ensure data consistency among different sources.