

SOMIA DOUSA

(315) 760-7703 • sabdelra@syr.edu • [linkedin.com/in/somia-dousa](https://www.linkedin.com/in/somia-dousa) • Portfolio

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY May 2024
M.S. Applied Data Science Fulbright scholarship | Convocation Master's Award recipient (Top of 200+ class) | GPA: 4.0
Relevant Coursework: Applied Machine Learning | Database Management | Tableau | Statistics | Data Warehouse
Petronas University of Technology, Engineering Department, Perak, Malaysia May 2016
B.E. Electrical and Electronics Engineering Petronas scholarship | Dean's List (2012-2016)

RELATED EXPERIENCE

Graduate Research Intern, School of Information Studies, Syracuse University Jul 2024 – Present

- Conducted self-paced research on building chat bots and agents using Python and Streamlit, including exploration of development environments and deployment of web applications.
- Developed chat bots with functionalities such as user interaction, short-term memory using LLMs, enhanced user interfaces, and prompt engineering, documenting best practices and challenges throughout.

Teaching Assistant, School of Information Studies, Syracuse University Aug 2022 – May 2024
Courses: Introduction to Front End Development | Database management

- Evaluated 120 student labs and assessments weekly and conducted tutoring sessions to enhance learning outcomes.

Data Analyst (Service Quality Management Specialist), Zain Telecommunication, Sudan Sept 2020 – Aug 2022

- Developed comprehensive weekly network analysis presentations and interactive dashboards for top management using SQL and Python Dash, leveraging data from a database of over 12 million customers to provide real-time insights and visualizations that improved strategic decision-making and operational oversight.
- Analyzed and managed datasets from telecom probing systems using SQL, handling over 12 million records to troubleshoot and escalate performance issues, resulting in improving the resolution time.
- Collaborated with marketing and B2B teams to conduct in-depth customer data analysis, supporting targeted complaint resolution and optimizing marketing campaigns, which resulted in a 25% increase in campaign effectiveness and customer satisfaction.
- Awarded Top Performer in the department in 2021 for exceptional contributions, including reducing customer service agent workload by 50% through data-driven insights and alternative communication channels.

PROJECTS

Predictive Modelling to Address Health Disparities in Cancer Diagnosis Jan – April 2024

- Utilized PySpark to develop and validate predictive models (Random Forest, Gradient Boost Classifier, Decision Tree, Logistic Regression) for forecasting metastatic cancer diagnosis timing, achieving an AUC of 81.35% with the Random Forest model.
- Processed and analyzed a dataset of 12,906 patient records using PySpark, performing feature selection and exploratory data analysis to assess the impact of demographic, socioeconomic, and environmental factors on diagnosis timing.
- Identified key factors contributing to late diagnoses through PySpark data transformations and machine learning pipelines, providing insights for healthcare policy improvements to reduce disparities in cancer diagnosis and treatment.

Environmental Impact Assessment and Policy Formulation for Flood Management Jan – April 2023

- Built a deep learning model for detecting green spaces in urban regions for future revitalization.
- Preprocessed the dataset by resizing images and normalizing pixel values.
- Constructed a CNN architecture, to classify images into categories such as buildings, streets, and green spaces(trees).
- Achieved 80% accuracy on a test set of labeled images, with a precision of 98% for the green spaces category.
- Utilized Python, TensorFlow, and OpenCV for data preprocessing, modeling, and visualization.

TECHNICAL SKILLS/CERTIFICATIONS

Programming Languages: Python (NumPy, Pandas, Scikit-learn), R, C/C++
Database Management: SQL- Microsoft SQL Server, Snowflake
Software: MATLAB, SPSS, Tableau, PowerBI
Statistical Analysis: T-test, Regression analysis, ANOVA, and chi-square tests.
Machine Learning: Regression, Classification, PCA, CNN

LEADERSHIP EXPERIENCE

Ambassador, Women in Data Science (WiDS), Cleveland, Ohio Jul 2024 – Present

- Lead the efforts to organize data science events to increase women representation in the field.