RM

Email
mutobvuronewa@gmail.com

Phone
079 545 6471

Location
Midrand
Willing to relocate

Demographics
Female
South African
Birthday: 19 July 1997

LinkedIn
linkedin.com/in/mutobvu-ronewa

Ronewa Ouma Mutobvu

I am a detail-oriented and innovative data scientist graduate with a solid educational background and experience in the banking sector. With over 2+ years working with Python, R, SAS, and SQL. I am deeply passionate about applying mathematical and statistical principles, as well as data-driven approaches, to address challenges and deliver impactful solutions. My skill set encompasses various areas including machine learning, data manipulation, data analysis, and problem-solving. My passion lies in leveraging these techniques to extract valuable insights from data and drive informed decision-making. Additionally, I possess strong leadership skills and thrive in collaborative environments. I have a strong drive for continuous learning and professional growth.

Experience

February 2023 – Present
Data Science (Intern) | Absa Group | Johannesburg

Worked in a Home Loan Non-take up text classification project: tasked to develop a model that can classify withdrawal comments into predefined reason groups and sub-groups, given free-text format dataset.

- Conducted extensive data preprocessing to clean and prepare the free-text withdrawal comments for analysis. Performed tokenization, lowercasing, and removal of stop words and special characters.
- Collaborated with domain experts to define the predefined reason groups for classifying the withdrawal comments. Annotated the dataset by manually reassigning labels to each withdrawal comment along with providing their descriptions and reasons.
- Relabeled any misclassified instances to improve the accuracy of the dataset.
- Explored various NLP algorithms and models, including but not limited to, Logistic Regression, Linear Support Vector Classifier, and deep learning- based approaches such as recurrent neural networks (RNNs) and transformer models like BERT.
- Extracted relevant features from the preprocessed text data to represent the comments effectively for the classification task. Experimented with word embeddings, TF-IDF vectors, and other representations.
- Trained the selected NLP model using the annotated dataset. Conducted extensive hyperparameter tuning and cross-validation to optimize model performance.
- Evaluated the model's performance using appropriate metrics such as accuracy, precision, recall, and F1-score. Conducted error analysis to identify areas for improvement.
- Implemented techniques to handle the imbalanced distribution of withdrawal comments among different reason groups.
- Built a user-friendly web application using Streamlit. The web application allows end-users to input withdrawal comments or browse in a file of comments and obtain the corresponding reason group classification from the trained NLP model in real-time.
- Practical training on binning of propensity scores and analyzing exclusions and overdrafts in the context of personal loans using SAS.

September 2022 – December 2022
Data Engineer, Record Management, Data Privacy (Internship) |
Explore AI Academy in partnership with FNB | Johannesburg

- Designing, developing, and implementing ETL processes to integrate data from various sources into the data infrastructure. This involves understanding the data sources, mapping data elements, and ensuring data quality during the extraction, transformation, and loading stage.
- Managing the tables and files within the big data platform, including creating new tables, updating existing ones, and organizing files for efficient storage and retrieval.

- Consolidating and joining training and KPI list into one comprehensive excel document to ensure contractors/stakeholders are up to date with their training obligations.
- Demonstrating an understanding of cooperation dynamics and actively engaging with diverse individuals across different projects. This includes fostering effective communication, building relationships, and promoting collaboration to achieve project goals.
- Creating engaging training policy content to communicate training requirements, guidelines, and procedures to contractors and other stakeholders. This involve developing training videos, policy gamification, and documents that are easy to understand and follow.

January 2022 – August 2022
Data Science (Learnership) | Explore AI Academy | Johannesburg

- Collaborated with teams and managed projects.
- Developed advanced informative dashboards using Power BI.
- Managing data, analytical programming, machine learning, and model building.
- Used programming languages - such as Python and SQL - clean, analyze, and interpret data.
- Designed a Streamlit web app.

Education

Jan 2021 - Dec 2022
Integrated Masters' Degree | Data Science/ E- science
University of Witwatersrand & University of Limpopo, South Africa

- (Wits) Capstone project: Differential gene expression in schizophrenia: A secondary analysis of public RNA-seq datasets using R-language,
- (UL) Thesis: Prediction of South Africa crime rate using supervised machine learning techniques.

Jan 2020 - Dec 2020
Honours | Statistical Science
University of Limpopo, South Africa

- Thesis: an impact of modularisation on the grade 12's pass rate in Elim area of Limpopo province, South Africa.

Jan 2016 - Dec 2019
Bachelors' Degree | Mathematics & Statistical Science
University of Limpopo, South Africa

Key Projects

South African Language Identification Classification
NLP Hackathon
- Using Python and Kaggle.

Differential Gene Expression in Schizophrenia
Secondary analysis of public RNA-seq datasets
- Using R-language and ZA-Wits-Core (core cluster).

Load Shortfall
Predicting the shortfall between energy generated by fossil fuels and with renewable sources (Hackathon)
- Using Python and Flask webserver with the help of AWS.

Soccer Match Prediction using Machine Learning
Using Europe league dataset
- Using Python.

Climate Change Belief Analysis
Sentiment analysis
- Using Python, version control comet and open-source app framework Streamlit
- Collaboration on Kaggle and GitHub.
- Used supervised Machine Learning to analyze whether the tweets are neutral, positive, or negative about the climate change.
- Utilized Python to build the algorithm and Flask to deploy it.

Movie Recommender System (Unsupervised Machine Learning)
Used collaborative and content-based filtering (Hackathon)
- Using Python, version control on comet.
- Collaboration on GitHub and Kaggle.
- Designed and implemented movie recommendation applications using Python, NLP, and Unsupervised Machine Learning models.
- Utilized Streamlit App and AWS EC2 instance to deploy the movie recommender application.
- This system recommends the movies that are relevant to the user's preferences based on either the ratings they have given to other movies or the genre, actors, and producers of the movies they have previously watched.