Portfolio of Harriet Sibitenda

Contact Information:

- Email: nkarietn@gmail.com, hsibitenda@wpi.edu
- Phone: +1 508 410 18 30, +221 774816846

---

1. PhD Thesis Project: Analyzing Social Network Data to Identify Development Tracks for Africa

Description:
My PhD research explores the application of **Artificial Intelligence (AI)** and **Machine Learning (ML)**, including **Large Language Models (LLMs)**, to analyze social media data related to development issues in Africa. The project focuses on:

- **Data Collection**: Using APIs and web scraping tools (such as **YouTube Data API**, **Snscrape**) to gather large datasets of 22,036 records from platforms like Twitter and YouTube.
- **Topic Modeling**: Using **BERTopic** and other text clustering techniques to identify and categorize common themes of discussion. Extracted thematic topics from comments based on document similarity by LLMs text-based generated labels.
- **Sentiment Analysis**: Applying AI models (e.g., **Llama2**, **BERT**, **VADER**) to predict sentiment on development issues such as poverty, education, and governance.
- **Social Network Analysis**: Integrating **Graph Convolutional Networks (GCNs)** and entity recognition techniques to map relationships between influential actors and development topics.
- **Time Series Forecasting**: Using models like **ARIMA**, **LSTM**, and **Transformer** to predict trends in public discourse related to key development topics.

Key Results:

- We identified 304 distinct topics with a coherence score of 0.81, 98 topic-based themes and futher analyzed existence of six 6 themes (hunger, poverty, education, security, employment and health) in comments with an F1 score of 0.89 for poverty-related themes using **Llama2**.
- Successfully implemented sentiment analysis by text-based generation by Llama2 with 0.55 balanced accuracy and 0.72 precision.
- Applied **social network analysis (SNA)** to identify key influential social actors in sectors such as education and business.

---

2. Research Papers and Publications

- **Paper 1**: "Retrieving Data from Social Network Platforms" (Presented)
  Summary: This paper reviews methods for collecting data from social media platforms and

provides insights into the limitations and strengths of various tools like APIs and web scrapers.
Link: waiting

- **Paper 2**: "Extracting Semantic Topics about Development in Africa from Social Media" (Published)
Summary: This publication introduces advanced topic modeling techniques such as BERTopic to extract development-related themes from African social media discourse.
Link: https://ieeexplore.ieee.org/document/10689600

- **Paper 3**: "Leveraging LLMs for Integrated Sentiment and Topic Analysis on African Social Media" (Accepted)
Summary: This work demonstrates the integration of **LLMs** for sentiment and topic analysis, focusing on their application in African social media to identify public concerns.
Link: waiting

- **Paper 4**: "Temporal Analysis of Social Concerns on African Social Media: Insights from Topics, Themes, and Sentiments" (Submitted)
Summary: This work discusses the trends of revealed social concerns by topic modeling, thematic and sentiment analysis using time series forecasting

  Link: waiting

---

3. Technical Projects

- **Project Name**: **Social Media Data Analysis for African Development**
Objective: Built an AI-driven framework to monitor and analyze social media discussions on development topics on data retrieved from social networks of Twitter and YouTube

  Technologies Used: Python, Selenium, Beautifulsoup, Snscrape, BERTopic, **Llama2**-3, **BERT**, Bart-REBEL, Graph Convolutional Networks (GCNs).
Outcome: Identified key development trends and influential actors in public discourse across Africa, supporting data-driven policymaking.
Link to Code/Repository: https://github.com/SibitendaHarriet/Data_collection_common_topics

- **Project Name: Smart Ad A/B Testing**

  Objective: Conducted sequential A/B testing to evaluate the impact of advertisements on user behavior by integrating machine learning for hypothesis testing.

  Technologies Used: Python, Scikit-learn, T-test, Random Forest, Sequential A/B Testing.

  Outcome: Improved understanding of user interaction on the Smart Ads website, ensuring error levels below 5% in the testing process.

  Link to Code/Repository: https://github.com/SibitendaHarriet/abtest-mlops

- **Project Name: User Acquisition Analysis in the Telecommunication Industry**

Objective: Assessed the impact of a telecommunication company by analyzing user behavior from XDR sessions across various applications like social media, Google, and YouTube.

Technologies Used: Python, MySQL, Random Forest Classifier, Streamlit.

Outcome: Achieved an 84% accuracy rate with Random Forest Classifier for predicting user engagement duration. Deployed the model on Streamlit and integrated data into MySQL for back-end usage.

Link to Code/Repository: https://github.com/SibitendaHarriet/PythonPackageStructure

- **Project Name: African Language Speech Recognition**

Objective: Developed an Automatic Speech Recognition (ASR) model to transcribe speech in Swahili, enabling machines to understand and respond to verbal commands in African local languages.

Technologies Used: Python, Deep Learning, Neural Networks, Acoustic Modeling, MFCC, Flask for deployment.

Outcome: Successfully deployed the speech-to-text system with competitive word error rates for Swahili. Utilized DVC for data versioning and MLFlow for model tracking.

Link to Code/Repository: https://github.com/10Academy-Group-4/Week-4/tree/speechhary

- **Project Name: Pharmaceutical Sales Prediction for Rossman Pharmaceuticals**

Objective: Created a machine learning model to forecast product sales across multiple stores six weeks in advance.

Technologies Used: Python, Random Forest, Decision Tree, LSTM, Sklearn Pipelines.

Outcome: Achieved high predictive performance with LSTM models (RMSE = 7,520,946). Analyzed the relationship between sales, promotions, and special seasons.

Link to Code/Repository: https://github.com/SibitendaHarriet/sales

---

4. Presentations and Conference Talks

- **Presentation 1: Poster Presentation at WPI, Worcester, MA, USA**

Title Extracting Topics from Social Networks using Llama2

Event: Research, Discovery & Innovation (ReDI) Fourth Annual Symposium, April 2, 2024, at WPI

Date: March 2024

- **Presentation 2: Review of Data collection for Social Network Comments**

  Title: " Retrieving Data from Social Network Platform"

  Event: CNRIA2023, Senegal-Gambia

  Date: May 2023

- ***Presentation 3*: *Pannel discussion on ICT tools for Africa Sustainable development***

  Title: Potential of AI for Africa

  Event: PASET conference, Mohamed 6 Polytechnic Institute, Morocco

  Date: June 2022

---

5. Awards and Recognition

- **Scholarship Award**
  Institution: PASET-RSIF
  Date: June 2020
  Description: Awarded for PhD research project

---

6. Conclusion

My journey through Artificial Intelligence (AI), Machine Learning (ML), and social network data analysis has been deeply rewarding, and it has equipped me with a robust technical toolkit to address complex problems in real-world contexts. My focus on applying LLMs, topic modeling, sentiment analysis, and social network analysis to development issues in Africa has not only enhanced my research but also my ability to contribute meaningful insights to policymakers and social advocates.

As I move forward, I am excited to expand on my research and technical skills in educational technology and other interdisciplinary fields. I am passionate about bridging AI research with real-world applications, and I am eager to bring this expertise to future collaborations and innovative projects.

Thank you for taking the time to review my portfolio. I look forward to the opportunity to further discuss how my skills and experiences can contribute to future research endeavors and collaborative projects.
