

FAITH WAVINYA MUTINDA

Email: wavinya.faith@gmail.com ◇ Website: <https://faith-wm.github.io/index.html>◇

SUMMARY

Data Scientist with experience in research, development, and implementation of machine learning and deep learning models including classification, regression, and natural language processing (NLP). Experience in Python programming and packages such as Pytorch, Tensorflow, Scikit-learn, CUDA, and familiar with large language models.

EDUCATION

Nara Institute of Science and Technology <i>PhD in Information Science and Engineering</i>	October 2020 – March 2023 <i>Nara, Japan</i>
Nara Institute of Science and Technology <i>Master in Information Science and Engineering</i>	April 2019 – September 2020 <i>Nara, Japan</i>
Osaka University <i>Research Student (Big Data Engineering Laboratory)</i>	April 2017 – March 2019 <i>Osaka, Japan</i>
Kenyatta University <i>BSc Telecommunication and Information Technology</i>	May 2011 – December 2015 <i>Nairobi, Kenya</i>

WORK EXPERIENCE

Biofourmis Singapore PTE. Limited <i>Data Scientist (Research & Development)</i>	April 2023 – Present <i>Singapore</i>
<ul style="list-style-type: none">Developed predictive AI models aimed at identifying patients at high risk of chronic diseases, hospital readmissions, and mortality. Several models achieved over 90% accuracy ranking within the top models in the precisionFDA challenge.Researched and evaluated large language models such as BERT and ChatGPT for clinical NLP applications such as named entity recognition, question answering, medical coding, and automatic selection of patients eligible for clinical trials.	
MES (UK) Limited, Sesby <i>Support Staff</i>	July 2016 – March 2017 <i>Nairobi, Kenya</i>
<ul style="list-style-type: none">Supported the production and quality assurance team to ensure timely delivery of orders.Prepared summary reports on the quality status of products based on quality assurance team reports.	
Safaricom Limited <i>Business Solutions and Architecture Intern</i>	October 2015 – Jan 2016 <i>Nairobi, Kenya</i>
<ul style="list-style-type: none">Collaborated with field technicians and scheduled surveys to determine the best services to offer to clients.Collaborated with field technicians and sales teams to prepare competitive business and technical proposals for clients.	
Jamii Telecommunications Limited <i>Network Operations Center Intern</i>	May 2014 – August 2014 <i>Nairobi, Kenya</i>
<ul style="list-style-type: none">Coordinated with field technicians to ensure fast fault resolution within the slated service level agreement.	

- Coordinated with other network monitoring teams to minimize service outages and improve client satisfaction.

PROJECTS

Disease Risk Prediction

- Implemented machine learning-based models that leverage electronic health records (EHR) data to predict patients at high risk of chronic diseases (such as cardiovascular, pneumonia, diabetes, atrial fibrillation, chronic kidney diseases), hospital readmissions, and mortality. The cardiovascular risk prediction model was among the top models in precisionFDA Challenge.

Automatic Data Extraction to Support Meta-Analysis Statistical Analysis

- Developed a transformer-based deep-learning model to support automation of clinical domain meta-analysis process. The model automatically extracts core information from clinical trials research publications, converts the data into a structured format, performs statistical analysis, and visualizes the results.

Contextualized medication event extraction

- Participated in a shared task challenge and developed a deep-learning based model for extracting medication mentions from clinical notes and classifying the medications into different classes depending on the context. The model was among top 10 best models in the n2c2 2022 CMED shared task/challenge.

Semantic textual similarity in clinical domain texts

- Developed a deep-learning model for computing the degree of semantic similarity in English and Japanese clinical texts. Semantic textual similarity plays an important role in natural language processing applications such as plagiarism detection, text summarization, and information retrieval.

SKILLS

Programming Advanced (Python), Basic (Java, JavaScript, C++, PHP, Swift, SQL, R, React) | Natural Language Processing | Machine Learning | Deep Learning

SELECTED PUBLICATIONS

- **Mutinda, F.W.**, Liew K., Yada, S., Wakamiya, S., & Aramaki, E. (2022). Automatic Data Extraction to Support Meta-Analysis Statistical Analysis: A Case Study on Breast Cancer. BMC Medical Informatics and Decision Making.
- **Mutinda, F. W.**, Liew K., Yada, S., Wakamiya, S., & Aramaki, E. (2022). PICO Corpus: A Publicly Available Corpus to Support Automatic Data Extraction from Biomedical Literature. In Proceedings of the First Workshop on Information Extraction from Scientific Publications. Asia-Pacific Chapter of the Association for Computational Linguistics.
- **Mutinda, F.W.**, Yada, S., Wakamiya, S., & Aramaki, E. (2022). AUTOMETA: Automatic Meta-Analysis System Employing Natural Language Processing. MEDINFO 2021: One World, One Health-Global Partnership for Digital Innovation.
- **Mutinda, F.W.**, Yada, S., Wakamiya, S., & Aramaki, E. (2021). Semantic Textual Similarity in Japanese Clinical Domain Texts Using BERT. Methods of Information in Medicine.
- **Mutinda, F.W.**, Nigo, S., Shibata, D., Yada, S., Wakamiya, S., & Aramaki, E. (2020). Detecting Redundancy in Electronic Medical Records Using Clinical BERT. The Association for Natural Language Processing.