Michael Geraldin Wijaya

Jakarta, Indonesia | +62-895-6043-39001 | michaelwjy2@gmail.com
LinkedIn | GitHub | Portfolio

EXPERIENCE

Tutor Bina Nusantara University

Bina Nusantara University 09/2023 – 01/2024

- Conducted weekly study classes for Statistics and Linear Algebra, supporting students throughout one semester.
- Facilitated learning by providing explanations, answering questions, and offering guidance to ensure student success in these subjects.

Project-Based Virtual Intern: Data Scientist – ID/X Partners x Rakamin Academy

ID/X Partners, Jakarta, Indonesia 08/2023 – 09/2023

- Finishing various task related with the activity of Data Scientist from ID/X Partners, such as Big Data Fundamental, Statistics & Data Analytics, and Machine Learning Implementation.

EDUCATION

Bina Nusantara University
Bachelor of Computer Science | 2022–Present
Streaming: Intelligent System
Latest GPA: 3.94
Achievements:
- Awardee Widia Partial Scholarship BINUS University
- Awardee of Beasiswa Unggulan Kemendikbudristek 2023
- Published paper: *Comparative Analysis of Machine Learning Algorithms and Data Balancing Techniques for Credit Card Fraud Detection*, presented at the International Conference on Computer Science and Computational Intelligence (ICCSCI), and published in *Procedia Computer Science* (Elsevier)
- National Finalist ASEAN Data Science Explorers 2024 Competition
- Top 20 Finalist School of Computer Science Hackaton 2024 BINUS x Microsoft (AI4A)

Senior High School 4 Palembang
Natural Science | 2019–2022, Grade: 97.29
Achievements:
- 1st Place Winner of the National Independent Science Competition - Mathematics
- A+ (Gold Medalist) in the Indonesian Science Competition 2022 - Mathematics
- A+ (Gold Medalist) in the Indonesian Science Competition 2022 - Technology and Information
- Gold Medalist in the 2022 National Numeracy Olympiad for High School Students
- Gold Medalist in the National Student Science Competition - Mathematics

PROJECTS

Plant Disease Detection 10/2024 – Present

- This plant disease detection system utilizes traditional ML methods like SVM and Random Forest with feature extraction techniques (Hu Moments, Haralick textures, and Color histograms) and pretrained CNNs such as ResNet and MobileNet. It includes a FastAPI backend with TensorFlow Serving. The model is deployed on Google Cloud Platform (GCP) using Cloud Functions and Cloud Storage, with an additional Gradio interface for easy access on Hugging Face Spaces.

Reachout Mail Generator_GenAI_LLM 11/2024

- Outreach Mail Generator is a tool that helps companies to generate outreach emails using Groq, LangChain, Llama3.1, and Streamlit/Gradio. It allows users to input the URL of a company's careers page, extract job postings from it, and create personalized emails. These emails include relevant portfolio links retrieved from a vector database, tailored to match specific job descriptions.

PySpark ML 11/2024

- This PySpark_ML repository focuses on data processing and machine learning, covering PySpark concepts like data manipulation, transformation, and aggregation. It includes machine learning techniques using Spark's MLlib and ML libraries, with implementations of various regression and classification algorithms (e.g., random forest regressor, linear regression, logistic regression, naïve bayes, gradient boosting, and random forest classifier).

Movie Recommendation System_Coursera Machine Learning Specialization 11/2024

- These are assignments from a course in the Coursera Machine Learning Specialization, focused on implementing recommender systems using collaborative filtering and content-based filtering techniques like matrix factorization and neural networks.

Personal Medical Chatbot with LLM RAG 06/2024 – 07/2024

- Personal medical chatbot project using Large Language Models (LLM) and Retrieval Augmented Generation (RAG). The chatbot provides detailed answers by leveraging medical knowledge from PDF documents, employing LlamaCPP for LLM integration and llama_index for efficient document handling. This approach enhances natural language responses and enables advanced retrieval of health-related information.

Patient's Condition Classification Using Drug Reviews 06/2024

- The project uses drug reviews with machine learning models such as Passive Aggressive Classifier, Random Forest, and XGBoost to predict patient conditions. It applies text processing techniques such as CountVectorizer, TFIDFVectorizer, and Word2Vec to analyze user reviews. The project provides visualizations of condition proportions and word clouds, and recommends the top three drugs for the condition.

Simple Personal Chatbot 06/2024

- This simple personal chatbot project uses NLP techniques and neural networks to process user queries effectively. It employs TF-IDF and Word2Vec to enhance comprehension and responsiveness. Data preprocessing techniques like tokenization, punctuation removal, and stop word elimination help the chatbot learn from an intents.json dataset to categorize patterns and responses. TF-IDF evaluates word importance, while Word2Vec creates word vectors for better semantic understanding. Flask is used to develop API endpoints, enabling integration with various platforms.

Ad Hoc Insights Consumer Goods 06/2024

- This project aimed to analyze consumer goods data to provide insights for management decisions. Using MySQL, I implemented concepts such as Joins, Subqueries, String, Numeric, and Date Functions, as well as Window Functions and Common Table Expressions (CTEs) for effective data manipulation. To visualize the findings, I used Tableau to create clear and informative data visualizations, and Canva for presentation creation.

LibraLink: Smart Attendance with Book Recommendation 09/2023 – 12/2023

- LibraLink is an advanced attendance system integrated with a book recommendation feature, specifically designed for library environments. At its core lies a robust facial recognition-based attendance mechanism, powered by Convolutional Neural Networks (CNN). In addition to tracking attendance, the system offers a personalized book recommendation service, leveraging the K-Nearest Neighbors (KNN) algorithm to align with user preferences. The system uses MySQL as its relational database to manage and store student data.

Movie Recommendation System 08/2023

- This ML project utilizes a Content-Based Recommender System for movies, using features like titles, genres, and posters to provide recommendations based on the chosen film. The system employs cosine similarity to gauge similarities between movies and is deployed using the Streamlit platform, enabling user-friendly interaction.

Image Classification of Billionaires 08/2023

- Utilized OpenCV for face and eyes detection, performed data cleaning through OpenCV's face detection, and conducted feature engineering using wavelet transforms. Built models using SVM, logistic regression, and random forest, further refining them with GridSearchCV. Exported final model to a file and established a Python Flask server. Deployed the project on AWS EC2.

Credit Risk Modelling 08/2023

- Utilized Logistic Regression to identify loan default risks from historical loan data. Conducted data preprocessing, feature engineering, and optimized model performance via ROC curve analysis. Gained insights into credit default risk factors, including loan duration, employment length, and payment history.

CERTIFICATES AND CERTIFICATIONS

- **Building Transformer-Based Natural Language Processing Applications** | NVIDIA | 2024
- **ASEAN Data Science Explorers 2024 National Final Indonesia** | ASEAN Foundation | 2024
- **Machine Learning Specialization** | DeepLearning.AI | 2024
- **Unsupervised Learning, Recommenders, Reinforcement Learning** | DeepLearning.AI | 2024
- **Advanced Learning Algorithms** | DeepLearning.AI | 2024
- **Supervised Machine Learning: Regression and Classification** | DeepLearning.AI | 2024
- **Oral Presentation** | International Conference on Computer Science and Computational Intelligence (ISSCSI) | 2024
- **Data Science and Machine Learning Fundamentals** | Corporate Finance Institute (CFI) | 2022
- **Machine Learning with Python** | IBM | 2022
- **Statistics Fundamentals** | Corporate Finance Institute (CFI) | 2022
- **Data Analysis using Python** | IBM | 2022
- **Python for Data Science** | IBM | 2022

SKILLS

- Programming: Python, SQL, C, Java
- Data Visualization
- GenAI & LLM: LangChain, LlamaIndex, ChromaDB, FAISS

- Machine Learning: Machine Learning Algorithms and Techniques
- Deep Learning: Tensorflow, Keras
- Natural Language Processing (NLP)
- MLflow
- Git/GitHub/DagsHub
- Flask/FastAPI
- Docker
- Google Cloud Platform (GCP) Products: BigQuery, Cloud Storage, Cloud Functions, AWS: EC2
- PySpark

ORGANIZATIONS AND ACTIVITIES

- Data Science Club (DSC) and Ureeka BINUS University (2024 – Present)
- Badminton (2023 - Present) and Binusian Gaming (2022 – Present) BINUS University
- Tutor BINUS University (2023 – 2024)

VOLUNTEERING

- Anti Bullying Educator - Christian Eunike Orphanage Foundation, Semarang, Central Java (11/2022)
- Patriotism Educator - Bethlehem Salvation Army Orphanage, Semarang, Central Java (06/2023)