SITHMI WIJESINGHE

Minuwangoda, Sri Lanka

(+94) 76 457 1117 | wpehara20@gmail.com | linkedin.com/in/sithmiwijesinghe/

EDUCATION

University of Colombo – Faculty of Science *April 2021 - present*
B. Sc (Hons) in Data Science

GPA : 3.77/4 (First Class)

Recipient of the Dean's Lists for Academic Years : 2023/2024, 2021/ 2022, 2020/ 2021

Thesis Topic : Music Emotion Recognition for Sinhala & English Songs
-Grade : A
-Supervisor : Mr. E. R. A. D. Bandara

Musaeus College, Colombo 07 *2006 - 2019*
G.C.E. Advanced Level – Physical Science Stream *August 2019*

Results : Chemistry – A , Combined mathematics – B , Physics – B

Z-Score : 1.7041

DR(Colombo) : 331

G.C.E. Ordinary Level *December 2016*

Results : 9A's

WORK EXPERIENCE

OCTAVE- John Keells Group June 2025 – present
Intern Data Scientist

- Contributed to data analysis projects by performing exploratory data analysis using Databricks and Excel to support business insights. Assisted in maintaining and running dashboard codebases through PyCharm, ensuring smooth integration and functionality. Collaborated with delivery teams to translate analytical findings into actionable insights for business stakeholders.

Delmege Forsyth & Co. Ltd November 2024 – April 2025
Intern Data Scientist

- Actively contributing to the development of a machine learning module aimed at predicting demand for fast-moving consumer goods (FMCG). Collaborating with the IT department to collect and preprocess data, utilizing algorithms such as regression models and decision trees to analyze historical sales data and forecast future demand, and working on integrating the module into a company-wide application.

PROJECTS & PRESENTATIONS

Music Emotion Recognition for Sinhala and English Songs *June 2024 - March 2025*

Final Year Undergraduate Research

Designed and implemented a bilingual Music Emotion Recognition (MER) system to classify emotions in Sinhala and English songs using advanced machine learning and NLP techniques. The project addressed a major data gap in Sinhala music and explored the cultural challenges in emotion detection across languages.

Key Highlights:

- Developed a novel dataset of 3,728 songs (2,000 English, 1,728 Sinhala).
- Fine-tuned Emotion English DistilRoBERTa and XLM-RoBERTa models for automated emotion annotation using Ekman's six basic emotions.
- Achieved 58% accuracy (3-class model: happy, sad, anger) using XGBoost, SVM, and Random Forest after optimizing emotion categories.
- Identified cultural bias in audio features, where English features accounted for 70% of model importance, emphasizing the need for culture-specific modeling.

Contributions to the Field:

- Introduced large-scale bilingual MER dataset for Sinhala and English music.
- Demonstrated the importance of cross-cultural considerations in audio-based emotion recognition.
- Provided a baseline benchmark for future research in South Asian music emotion analysis.

Global Leptospirosis Surveillance and Predictive Analytics *August 2024*

Level 04 Group Project

Developed a global dashboard to track leptospirosis cases, with a special focus on Sri Lanka, to improve awareness and facilitate public health responses. The dashboard integrates and visualizes data from multiple sources, highlighting trends, patterns, and correlations in disease outbreaks. Time series forecasting models (AR, MA, ARMA, ARIMA) were applied to predict future case trends.

Contributions:

- Developed an R script to extract data from PDF tables in Sri Lanka's Epidemiology Reports.
- Designed and implemented a MongoDB database to store and manage extracted case data.
- Built the dashboard using Streamlit, focusing on Sri Lanka's leptospirosis trends.

Synergy HR Management *Spetember 2024*

Level 04 Group Project

Developed a predictive analytics system to identify key drivers of employee attrition and forecast employees at risk of leaving. Built multiple machine learning models, with AdaBoost yielding the highest performance. Explored neural networks with attention layers and hybrid models (voting and stacking classifiers). The final solution was integrated into a Flask-based web application with an interactive dashboard using Streamlit.

Contributions:

- Designed and developed the Flask web application to serve machine learning predictions.
- Integrated model APIs for real-time attrition risk assessment.
- Ensured seamless deployment and backend optimization for efficient request handling.

Unlock Tomorrow *August 2024*

Level 04 Group Project

Developed a machine learning-based prediction system to analyze and forecast term deposit subscription rates. The dataset exhibited two distinct clusters, leading to the development of separate models for each cluster. Addressed class imbalance using SMOTE, ADASYN, and random oversampling. Random Forest emerged as the best-performing model, achieving high accuracy and AUC-ROC scores while minimizing overfitting.
Contributions:

- Developed the machine learning model for one of the two clusters using advanced classification techniques.
- Performed outlier detection to enhance model robustness and data integrity.

Riding the Surge *February 2024*

Level 03 Group Project

A machine learning project focused on developing a dynamic pricing model for a ride-sharing company using historical ride data. The objective was to predict optimal ride fares and changes in profit percentage when shifting from a static to a dynamic pricing strategy. By evaluating multiple regression techniques, including linear models and tree-based methods, the XGBoost Regressor emerged as the best-performing model. A web-based data product was also developed, enabling real-time predictions of adjusted costs and profit margins based on key ride features

Regression Analysis for Health Insurance *2023*

Level 03 Group Project

Created a MLR model to predict the healthcare insurance costs and did a residual analysis. Contributed for the report.

Spotify and You Tube *September 2023*

Level 03 Group Project

Explored the characteristics of top-performing songs on Spotify and YouTube to identify trends in audience engagement and preferences. Analyzed features such as valence, tempo, energy, and album type, and correlated them with likes and stream counts to understand the elements that drive song popularity.
Contributions:

- Contributed to the development of interactive dashboards to visualize key insights related to song features and audience engagement.

SKILLS

Programming & WebDevelopment :Python, R,JavaScript,HTML/CSS,SQL,MongoDB, Flask, Streamlit

Tools & Technologies : Databricks, PySpark, PyCharm, LATEX(Overleaf),GitHub, R Studio, Visual Studio Code, Google Colab, Minitab, MS Excel, Power BI, Tableau

SoftSkills : Time & Stress Management,Teamwork,Attention to Detail,Creative Thinking, Problem Solving

EXTRA-CURRICULAR

Member of the Stat Circle of University of Colombo *2023- 2024*
Member of the organizing committee of Inter- University Quiz Competition, ML Nexus 2023

Tennis *2021 - 2022*
University Level

- 1st Runners Up (Team): Deuced Invitational Tennis Tournament 2022
- 2nd Runners Up: Inter-Faculty Tennis Tournament 2023

REFEREES

Dr. C. H. Magalla, PhD
Senior Lecturer
Department of Statistics
University of Colombo
Email: champa@stat.cmb.ac.lk

Mr. E. R. Anjana Damendra Bandara, MPhil
Senior Lecturer
Department of Statistics
University of Colombo
Email: anjana@stat.cmb.ac.lk

Dr. A. A. Sunethra, PhD
Senior Lecturer
Department of Statistics
University of Colombo
Email: sunethra@stat.cmb.ac.lk

Mr. Waruna Muwanwella
Assistant Manager
Group Analyst
Delmege Forsyth & Company
Phone: (+94) 713404438