PROJECT AND DATA MANAGEMENT PLAN. UKONU CHIZOBA MARYANN- 21089329

PROJECT TITLE: PREDICTIVE ANALYSIS ON MATERNAL HEALTH OUTCOMES USING IoT-DRIVEN DATA

RESEARCH QUESTION: "How can Deep hybrid models be optimized to achieve better performance than ensemble models in the classification of maternal health risks using physiological data?"

RESEARCH OBJECTIVE:

1. Train Machine Learning Models: Skillfully select and train various machine learning models to effectively classify maternal health risks.
2. Evaluate Model performance: Rigorously assess the model performance across different metrics to understand its general effectiveness.
3. Feature Importance Analysis: Analysis of the importance of each physiological feature in the selected classification models.
4. Build a deep hybrid model: Apply feature selection techniques to confidently incorporate significant features into a deep hybrid model.

PROJECT SUMMARY AND BACKGROUND

Maternal health encompasses the physical, mental, and social well-being of women during pregnancy, childbirth, and postnatal period. Maternal health is an integral component of a nation's wealth (M. Buckner, 2010). It is a main determinant of infant health but remains generally poor (F. Mavin et al., 1984). Pregnancy comes with its own set of potential risks. However, with the right prenatal care and support, the risk can be minimized and managed. The complexities surrounding the concept of maternal risk in prenatal care present challenges in accurately identifying high-risk women and providing corresponding care (Winikoff, 1995). A risk assessment approach to pregnant women has the potential to substantially decrease maternal mortality or morbidity (W. Phuapradit et al., 1990).

Developing a maternal risk model is a complex task, and its effectiveness is influenced by various factors. The utilization of IoT in healthcare shows promise for improving maternal health by continuously monitoring physiological parameters and providing real-time data for early detection and intervention. This project combines the strength of IoT technology and machine learning models to classify maternal health risk levels, particularly focusing on a hybrid model approach to enhance the accuracy and reliability of maternal health risk predictions. This comprehensive approach, barring socio-economic factors and underlying health conditions, aims to enhance healthcare outcomes for expectant mothers by providing precise risk assessments and supporting better healthcare planning and intervention.

REFERENCE

Ahmed, M., Kashem, M.A., Rahman, M., & Khatun, S. (2020). Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT). Available at: https://api.semanticscholar.org/CorpusID:214577407 (Accessed: 29 May 2024)

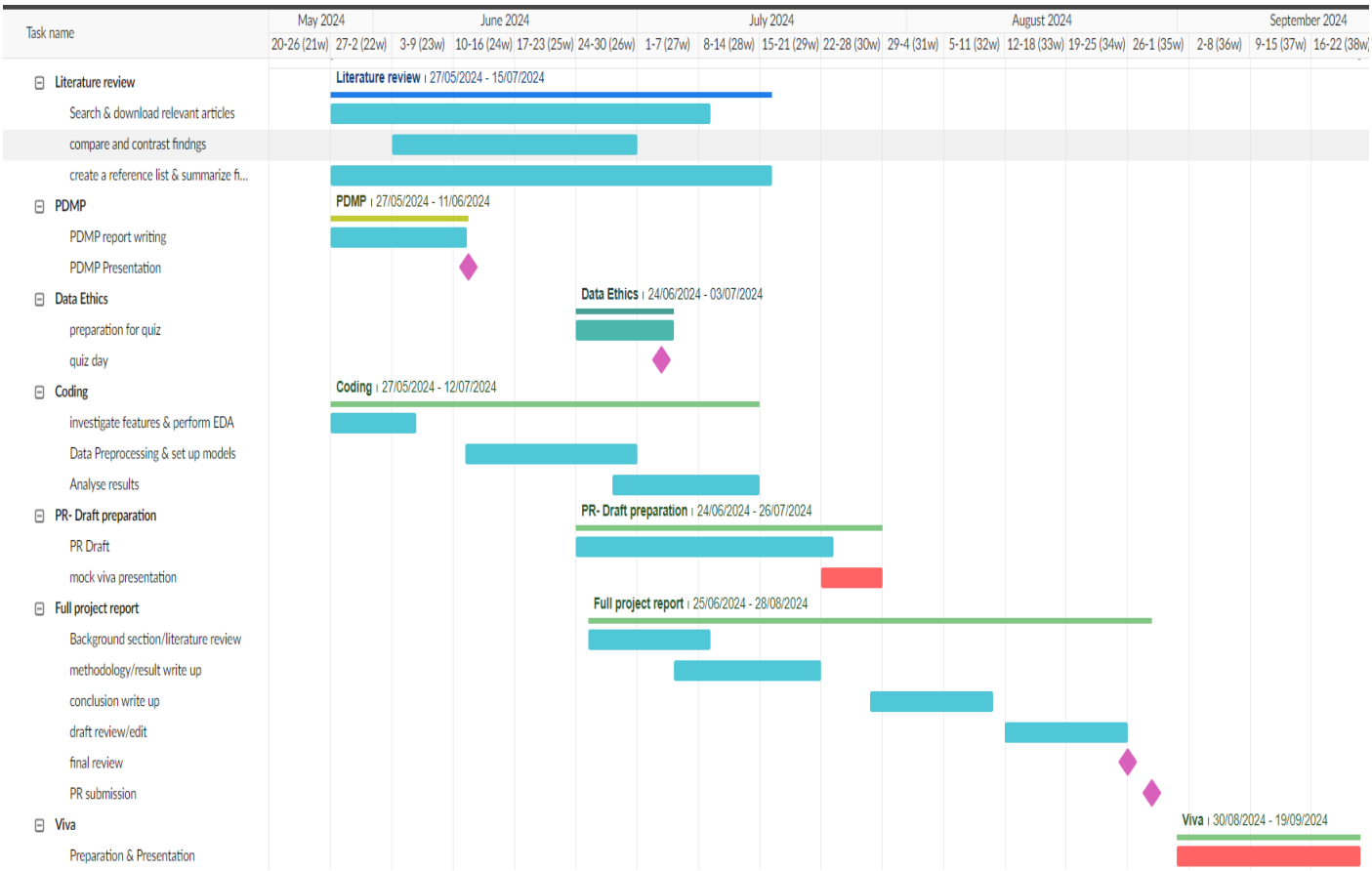
Buckner, M. (2010) 'Mothering the World: Improving Global Maternal Health', *Journal of Perinatal & Neonatal Nursing*. Available at: https://doi.org/10.1097/JPN.0b013e3181cfcaf1 (Accessed: 3 June 2024)

Favin, M.N., Bradford, B. and Cebula, D. (1984) 'Improving maternal health in developing countries'. Available at: https://api.semanticscholar.org/CorpusID:80475415 (Accessed: 1 June 2024).

Phuapradit, W., Pongthai, S., Sudhutoravut, S., Chaturachinda, K., & Benchakan (1990). Implementation of risk approach in maternal health. https://api.semanticscholar.org/CorpusID:80035094 (Accessed: 29 May, 2024)

PROJECT TIMELINE

TASK	DESCRIPTION	START DATE	END DATE
Literature Review	conduct an extensive research and review relevant articles on maternal health	1/6/2024	15/7/2024
Data set acquisition and supervisor approval	search for a suitable data set that meets all UH ethical considerations and obtain approval on its suitability on the project from the supervisor	17/5/2024	23/5/2024
PDMP	prepare and submit initial project and data management plan	29/5/2024	11/5/2024
Data Analysis/feature engineering/data preprocessing	Analyse the dataset using visualization/perform feature engineering where applicable	12/6/2024	16/6/2024
Data ethic quiz	prepare and write the data ethic quiz	24/6/2024	3/7/2024
Initial model set up and training	develop and train the models on the data set	17/6/2024	5/7/2024
Comparative Analysis of model performance	compare and contrast the results obtained by the trained models using suitable evaluation metrics	8/7/2024	12/7/2024
Model result documentation	Document the results and findings of the model and prepare report	13/7/2024	15/7/2024
Draft Project report	Prepare a draft report of my project for a mock viva presentation	1/7/2024	29/7/2024
Final Project Report	Bring together all my work including my literature review, model development, model evaluation and documentations of model findings and submit my report	8/7/2024	28/8/2024
Viva Presentation	Prepare for my viva	29/8/2024	19/9/2024



DATA MANAGEMENT PLAN

Data Collection: The dataset for this project was obtained from the University of California Irvine(UCI) Machine Learning Repository, specifically the Maternal Health Risk Data Set. The dataset can be accessed at the following link: Maternal Health Risk Data Set

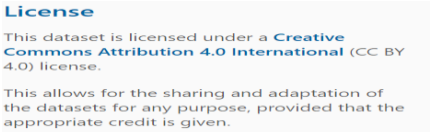
Overview of the Dataset: The dataset was created by: Marzia Ahmed (A. Marzia et al, 2020) of Daffodil International University. The data was collected from different hospitals, community clinics, and maternal health care in the rural areas of Bangladesh. The data was originated using IoT devices and manual recording. IoT is a wearable sensing-enabled technology that can continuously monitor physiological parameters like blood pressure, heart rate, blood sugar levels, and body temperature. The manual recording in this data was that of the age of the participants.

Summary of Data: The data is provided in a CSV(Comma-Separated Values) format, which is widely compatible with data analysis tools. It contains 1,014 records which spans over 6 features, each representing an individual with various physiological health indicators. The dataset size is approximately 36KB, making it manageable for analysis without requiring extensive computational resources. The code files such as jupyter notebook or Python script for the data processing and analysis are expected to be minimal in size.

How the Data Meets Ethical Requirements: GDPR Compliance: the dataset meets the GDPR requirements for data protection as it has been anonymized to ensure that no personally identifiable information is included.

UH Ethical Policy: The dataset was not a survey or a questionnaire and does not require UH ethical approval and it conforms with the GDPR policies.

Permission to Use Data: The creators of the dataset have made the dataset publicly available for educational and research purposes as can be seen from the below screenshot



Document Control: The files will be named using clear and consistent naming conventions and descriptive names(e.g., MaternalHealthData.csv, MaternalHealthcode.py). The version control where applicable will be managed in a repository hosted on GitHub. The link to the the Github can be accessed via: https://github.com/Maryannrichard/Maternal_health_research

Metadata: A ReadMe file will be provided in the GitHub repository which will include a brief description of the project and dataset with an explanation of its features, clearly commented codes, a list of required libraries and tools for running the scripts, and my contacts for questions or issues.

Security and Storage: The data and its associated code will be stored in various directories, including but not limited to GitHub, OneDrive, Google Drive, external drives, and local systems with restricted access. Regular updates and backups will be maintained to ensure data safety and availability. When necessary, data will be shared with a supervisor or authorized personnel through secure controlled platforms or internal university networks such as email.