# ELECTRONIC ASSIGNMENT COVERSHEET

| | |
|---|---|
| **Student Number:** | 34445873, 34420571, 33744747 |
| Student Name: | Chew Jian Yue, May Me Khin Maung Soe, Zachary Jamie Chan |

| | |
|---|---|
| **Unit Code:** | ICT606 |
| Unit name: | Machine Learning |
| Enrolment mode: | External |
| Date: | 17 June 2023 |
| Assignment number: | 1 |
| Assignment name: | Project |
| Tutor: | Mr. Johnson Ang |

**Student's Declaration:**

1. Except where indicated, the work I am submitting in this assignment is my own work and has not been submitted for assessment in another unit.
2. This submission complies with Murdoch University's academic integrity commitments. I am aware that information about plagiarism and associated penalties can be found at http://our.murdoch.edu.au/Educational-technologies/Academic-integrity/. If I have any doubts or queries about this, I am further aware that I can contact my Unit Coordinator prior to submitting the assignment.
3. I acknowledge that the assessor of this assignment may, for the purpose of assessing this assignment:
   a. reproduce this assignment and provide a copy to another academic staff member; and/or
   b. submit a copy of this assignment to a plagiarism-checking service. This web-based service may retain a copy of this work for the sole purpose of subsequent plagiarism checking, but has a legal agreement with the University that it will not share or reproduce it in any form.
4. I have retained a copy of this assignment.
5. I will retain a copy of the notification of receipt of this assignment. If you have not received a receipt within three days, please check with your Unit Coordinator.

| |
|---|
| I am aware that I am making this declaration by submitting this document electronically and by using my Murdoch ID and password it is deemed equivalent to executing this declaration with my written signature. |
| **Optional Comments to Tutor:**<br>*e.g. If this is a group assignment, list group members here*<br>Group assignment of three members **(Group 5)**<br>1. Chew Jian Yue (34445873@student.murdoch.edu.au) 34445873<br>2. May Me Khin Maung Soe (34420571@student.murdoch.edu.au) 34420571<br>3. Zachary Jamie Chan (33744747@student.murdoch.edu.au) 33744747 |

*If you can, please insert this completed form into the body of* ***each*** *assignment you submit. Follow the instructions in the Unit Information and Learning Guide about how to submit your file(s) and how to name them, so the Unit Coordinator knows whose work it is.*

***Start your assignment on the next page.***

# Table of Contents

## Abstract

Especially recently, there has been an increasing need for explainable machine learning (XML) techniques to be applied to assist in models predictions' understandability [1]. Increased emphasis is placed on the explainability, interpretability and trustworthiness of machine learning (ML) models, especially in the field of healthcare.

## Introduction

Perinatal mental health problems occur during pregnancy or during the first year after a child's birth [2]. The commonality of perinatal mental health issues, including perinatal depression, highlights the importance of primary care through early detection, treating and providing referrals to affected mothers [3].

As perinatal mental health problem continues to be a public health concern in this world [4], it is imperative to emphasize continuously in this field of research and determining the possible solutions via ML to lessen the risk factors relating to perinatal complications [5].

Possible risk factors include:

- Previous history of depression
- Life events
- Interpersonal conflicts

Poor mental health, like perinatal depression, is correlated with:

- Maternal, child and family challenges. Infants with perinatally depressed mothers have bad temperaments with impeded cognitive and emotional development [3].
- Increased infant morbidity and mortality rate [6].

Cost burden to governments on health and social care [6].

With early and easy identification of mothers with higher risk of perinatal complexity based on risk factors and features will lower the number of cases with perinatal complexity.

## Small-scale literature review

1. <u>Interpretability of predictive ML models</u> [7]

Descriptions of Decision Trees (DT), Regression Models (RM), and Naïve Bayes classifier as ML techniques approach towards interpretability. Introduces a different perspective to classifying approaches to ML interpretability as model-specific, model-agnostic and local, global.

| **Pros** | **Cons** |
|---|---|
| • Offers different perspectives to classification of interpretability in ML models.<br>• Describes the different classification methods of ML interpretability<br>• Recognises the importance of model interpretability for healthcare prediction models as opposed to other fields, largely because of needing to gain trust and confidence of healthcare practitioners | • Does not go into detail on simplifying complex models for explainability, possibly leading to sub-optimal results<br>• Existing methods to evaluate models' explainability, such as segmentation maps are poorly scalable.<br>• Paper emphasises more about interpretability in neural networks<br>• Introduction of new perspective to classification of ML interpretability approaches share minor and similar ideas as previous methods of classification (for example, model-agnostic methods are usually applied post-hoc) [8] |

2. <u>Perinatal health predictors using artificial intelligence</u> [4]

As ML has been developed to be provide early detection, monitoring of perinatal health as well as to predict possible contributing factors to perinatal health including "preterm birth, birthweight, preeclampsia, mortality and postpartum depression". ML used in this article can include supervised, semi-supervised and unsupervised learning such as: DT, Support Vector Machine (SVM), Logistic Regression (LR) and Neural Networks (NN) are included in this article.

| **Pros** | **Cons** |
|---|---|
| • ML models could provide better performance than traditional statistical models due to their capability to manage better with data that are nonlinear, numerous involvements between variables and handling of different predictors in any circumstances.<br>• ML can provide real-time predictions regarding the need for labour induction or caesarean delivery.<br>• Emphasises the need for development of tools to responsibly evaluate AI models for explaining its decision-making process to encourage social acceptance and integration<br>• To achieve the goal of explainability, collaboration with expert health clinicians is crucial | • Limitations include the possibility of providing inaccurate measurements and missing data.<br>• Ensure that ML developed is transparent and easily comprehended by humans which in this case, mainly clinicians otherwise would not be efficient if it is too complicated to use.<br>• No mention or suggestion on the feasibility of engagement with multiple facets of society from politicians to community partners |

## 3. Using ML to predict complications in pregnancy [6]

The article explains some possible risks that contributes to perinatal complications. Therefore, ML has been developed to predict whether if pregnant women will have an adverse result after giving birth. In this case, the systematic review was implemented using meta-analysis "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) and the interpretable ML models involved in this study were Random Forest (RF), LR, NN and SVM.

| **Pros** | **Cons** |
|---|---|

| | |
|---|---|
| • Systematic review indicating that 25% of studies utilised AI-interpretable ML models<br>• Allows to determine the most critical variable to prevent the adverse outcome from occurring<br>• Able to improve women's healthcare through providing diagnosis with higher accuracy, lessening physicians' workload and the healthcare costs as well as standardize analysis with concise interpretation. | • Although ML can help to enhance health care tremendously, there are cons and limitations of AI when used in health, including the necessity to consider ethical dilemmas as well as the possibilities of human biases when developing computer algorithms.<br>• When predicting in relations to Healthcare, this is dependent on races, genetics, gender and other attributes and traits otherwise could result in outcome being overestimated or underestimated of the patients' risks factors |

## 4. ML techniques for predicting depression and anxiety in perinatal women during COVID-19 [7]

In this article, ML algorithms were used to predict maternal depression and anxiety faced by females during the Covid-19 situations. Generally, Gradient Boosting (GB) and RF have the best prediction values on depression and anxiety. Furthermore, importance of contributing factors on depression and anxiety are evaluated and the most significant factors are mainly financial issues and stress during pregnancy.

| **Pros** | **Cons** |
|---|---|
| • Peer reviewed paper identified that results' interpretations are appropriate.<br>• Efficient in predicting maternal mental health using GB and RF. To be merged with clinical medical information for utilization | • Due to many possible key variables that could contribute to depression and anxiety, this could be very complex and complicated to work on. |

| | |
|---|---|
| • Increased risk of depression and anxiety are explained by financial problems (Q137), poor healthcare systems | |

## 5. Development and validation of ML algorithms for predicting risk of postpartum depression of perinatal women [9]

Similar to above articles, Postpartum depression (PPD) can be a life-threatening mental condition that could occur due to childbirth. To predict the risk of postpartum depression, interpretable ML algorithms including LR, RF and DT were integrated before implementation.

| **Pros** | **Cons** |
|---|---|
| • Algorithms were efficient in determining pregnant women with high potential of postpartum depression.<br>• Considerable dataset of 15,197 between 2015-2018 used | • Study limited to academic medical centres and does not reflect general overall US population. |

## Other insights and challenges to XML

DARPA defines explainability as models helping users understand and trust decisions of artificial systems [10]. As explained by [11], ML used in healthcare should be fair, secure, trustworthy, responsible, interpretable and explainable. However, there are challenges such as lack of formal definitions, and inconsistencies of explanation of model structure (e.g., feature importance). The guidelines for requirements of XML model are generic. Medical data are usually multi-dimensional, NN could be applied but explainability is lost, hence we have used mostly inherently explainable techniques (preferred to some clinicians).

# Experimental design

## Model Selection

The goal of predicting binary classes of Kessler K5+2 score to assess mental health risk in perinatal mothers defined as:

> "Low_risk" (0) and "High_risk" (1) classes indicate perinatal mother risk of experiencing mental health issues during perinatal period.

The dataset contains features relating to common life events, psychological distress, family violence, smoking, substance use, and challenges associated with stages of change. Accurate mental health risk prediction in perinatal mothers enables targeted support, personalized antenatal plans, and mitigates potential negative impacts.

We have selected three ML algorithms mainly LRs, DTs and RFs for this design as they are commonly used for binary classification tasks. Here's an explanation of why these algorithms are considered:

| **LRs** | **DTs** | **RFs** |
|---|---|---|
| • Suitable algorithm when dealing with binary classification problems.<br>• Used to estimate the probability of a perinatal mother being classified as high risk or low risk based on the provided features.<br>• Interpretable and provides insights into the significance of each feature's | • Intuitive and easy to understand, useful for generating interpretable rules in the classification process.<br>• Handles both numerical and categorical features.<br>• Recursively partitions the feature space based on the provided features, decision trees can classify perinatal | • Ensemble learning method that combines multiple DT to improve predictive performance<br>• Mitigates the risk of overfitting by aggregating predictions from multiple trees<br>• Provides feature importance rankings, which can help identify the most |

| | | |
|---|---|---|
| contribution to the prediction | mothers into high risk or low risk categories. | influential features in predicting the mental health risk of perinatal mothers. |

## Data Pre-Preparation

### Importing the data

We converted the dataset file to CSV and imported it. The converted csv dataset contains 24 columns, where the first 23 columns (*Q214-Q195*) except *participant number* represent the input attributes and the last column (Class) represents the output attributes.

### Missing values

There are no missing values, hence no need for data imputation.

### Renaming columns

Columns are renamed for user to easily recognise the variables. The naming conversion is based on the Excel file of the dataset.

### Dropping column

The "Participant Number" column in the dataset is not essential for analysis or classification tasks. The remaining dataset comprises the relevant features and the target variable, "Risk" column.

### Mapping output column

To use prediction models, it is necessary to convert the output variable from object type to numeric type and map its values to 0 and 1. To perform binary classification, we need to assign numerical values to these categories.

We encode "Low_Risk" to 0 and "High_Risk" to 1 allowing us to represent the two categories as numeric values, compatible with ML models.

## Encoding class labels

In the given dataset, the categorical variables are already represented as integer values instead of strings. However, for the purpose of avoiding any technical issues and ensuring compatibility with ML algorithms, the class labels have been encoded as integers.

## Splitting the data

To develop and evaluate ML models for predicting the 'Risk' variable, we need to split the available dataset into training and testing sets. This is essential to assess the model's performance on unseen data. The dataset is divided into two subsets: a training set, which contains 80% of the data, and a testing set, which contains the remaining 20%.

## Model Evaluation

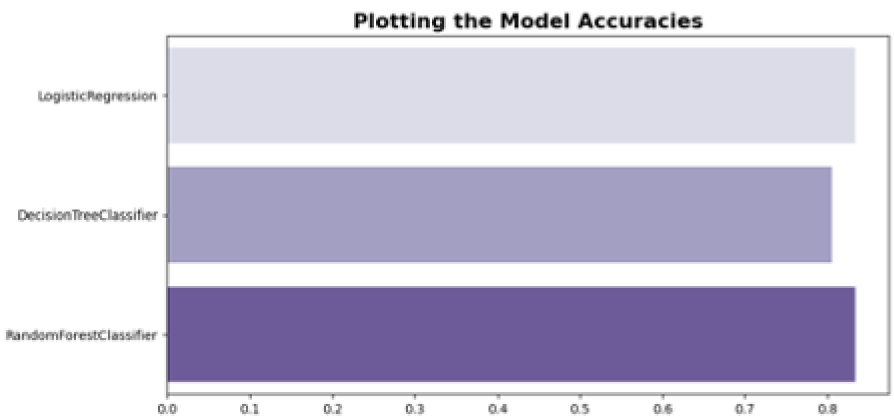
| **Model** | | **Precision** | **Recall** | **F1-score** | **Support** |
|---|---|---|---|---|---|
| **LR** | Low_risk 0 | 0.81 | 0.81 | 0.81 | 16 |
| | Hight_risk 1 | 0.85 | 0.85 | 0.85 | 20 |
| | Accuracy | | | 0.83 | 36 |
| | Macro Avg | 0.83 | 0.83 | 0.83 | 36 |
| | Weighted Avg | 0.83 | 0.83 | 0.83 | 36 |
| | | | | | |
| **DT** | Low_risk 0 | 0.74 | 0.88 | 0.80 | 16 |
| | Hight_risk 1 | 0.88 | 0.75 | 0.81 | 20 |
| | Accuracy | | | 0.81 | 36 |
| | Macro Avg | 0.81 | 0.81 | 0.81 | 36 |
| | Weighted Avg | 0.82 | 0.81 | 0.81 | 36 |
| | | | | | |
| **RF** | Low_risk 0 | 0.86 | 0.75 | 0.80 | 16 |
| | Hight_risk 1 | 0.82 | 0.90 | 0.86 | 20 |
| | Accuracy | | | 0.83 | 36 |
| | Macro Avg | 0.84 | 0.82 | 0.83 | 36 |
| | Weighted Avg | 0.84 | 0.83 | 0.83 | 36 |

The consolidated table includes several evaluation metrics: precision, recall, F1-score, and accuracy, providing insights on models' predictive capabilities and ability to classify instances of "Low_risk" and "High_risk" correctly.

LR: the precision and recall for both risk categories (0 and 1) are relatively high, suggesting that the model correctly identifies low and high-risk instances. The F1-score, which combines precision and recall, is also high for both categories.

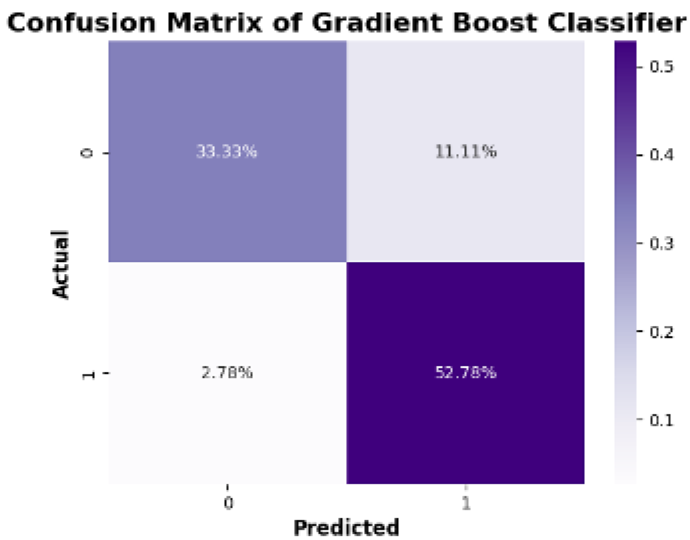
DT: the precision and recall for both risk categories are also reasonably high. The F1-score is relatively balanced for both categories, indicating a good trade-off between precision and recall.

RF is comparable to the other models. The precision and recall values are also high for both risk categories. The F1 score demonstrates a good balance between precision and recall for both categories.




For LR model, it achieved an accuracy of 0.83, correctly predicting the risk category 83% of the instances in test dataset. DT model achieved an accuracy of 0.81. RF classifier achieved an accuracy of 0.83.

All three models show promising performance in predicting the 'Risk' variable. They demonstrate reasonably high accuracy, with balanced precision and recall values.




True Positive (TP): 33.3%

It means that 33.3% of the High_risk cases were correctly identified by the model.

False Negative (FN): 11.111%

It means that 11.111% of the High Risk cases were missed or misclassified by the model.

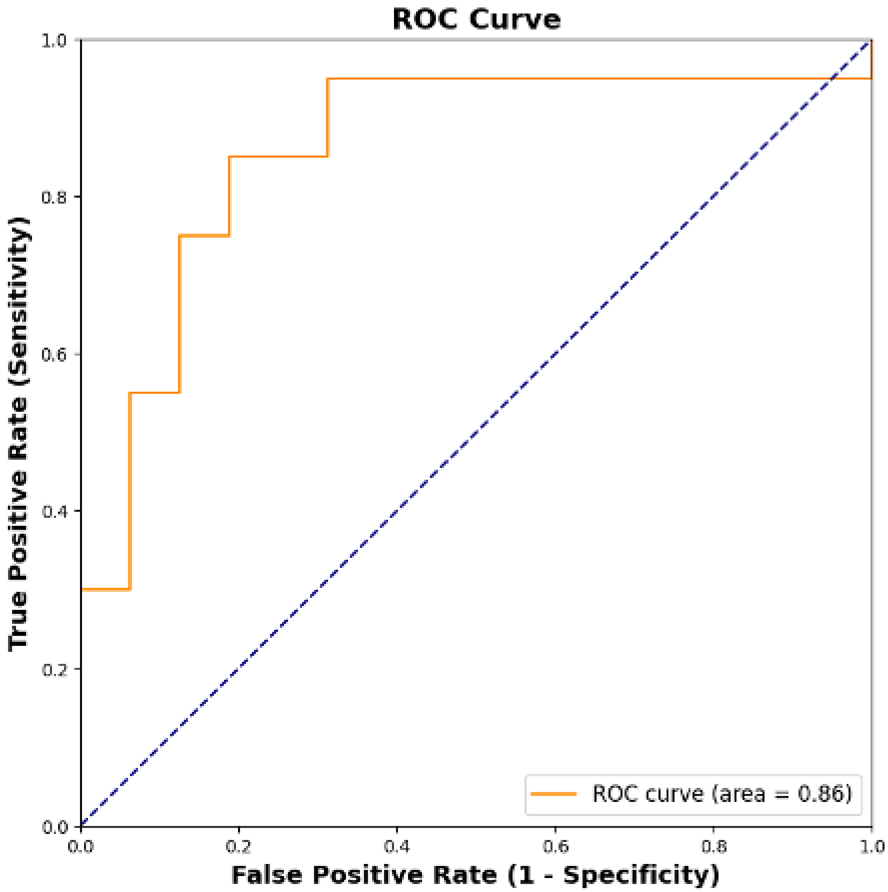
False Positive (FP): 2.78%

It means that 2.78% of the Low Risk cases were falsely identified as High Risk.

True Negative (TN): 52.78%

It means that 52.78% of the Low Risk cases were correctly identified by the model.

These results provide insights into the model's performance in distinguishing between “Low_risk” and “High_risk” instances. The values suggest that the model has relatively low false positive and false negative rates, meaning it performs reasonably well in predicting both classes.



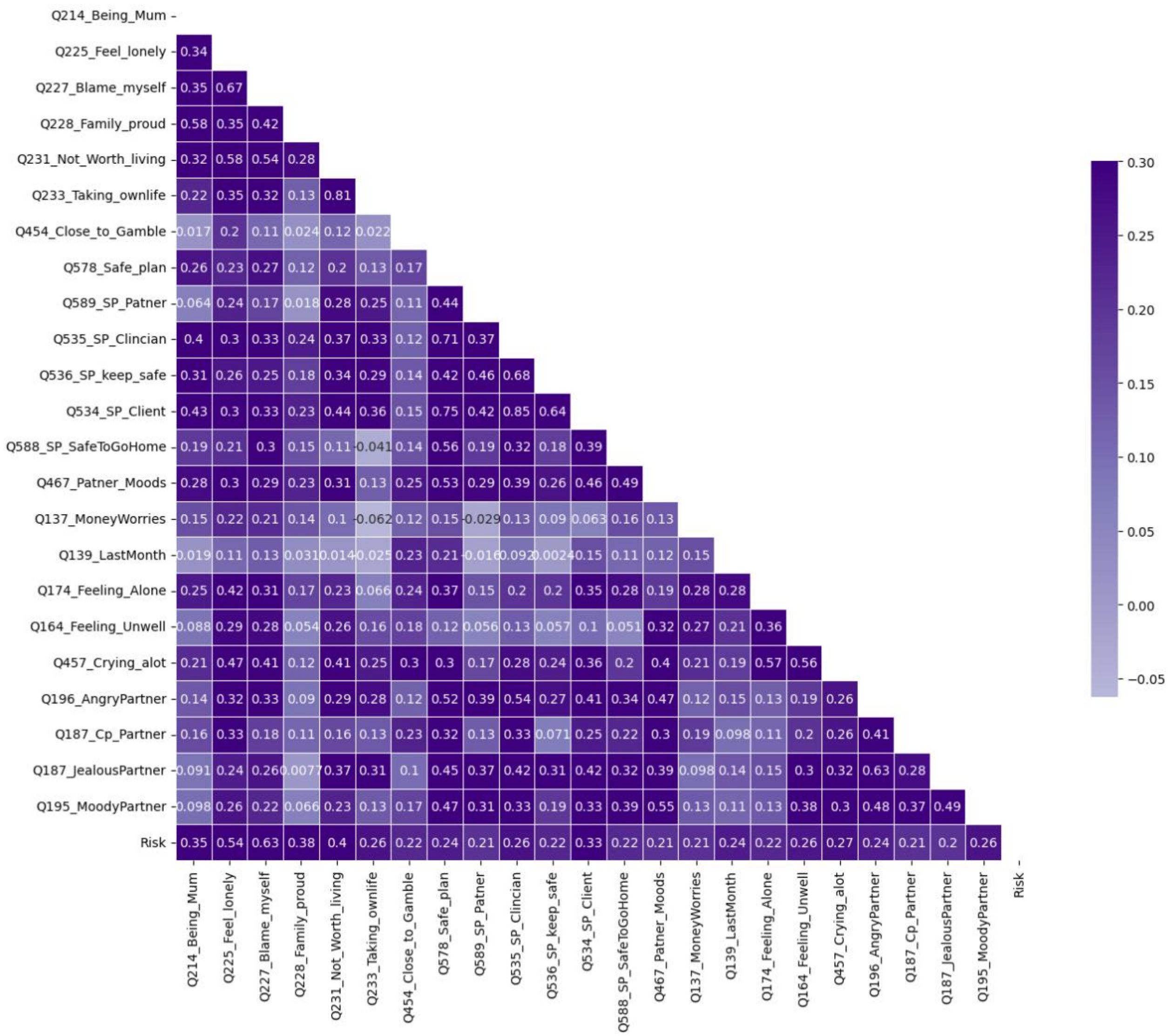

The true positive rate (TPR) vs the false positive rate (FPR) at various categorization criteria is shown graphically by the ROC curve. An indicator of a model's capacity to distinguish between positive and negative classes is the area under the ROC curve (AUC).

The model can efficiently distinguish between the positive and negative classes, according to an AUC of 0.86, indicating that it has good discriminatory power. The model's ability to correctly categorize cases improves with increasing AUC.

In conclusion, an AUC of 0.86 on the ROC curve suggests that the model performed rather well in terms of classification accuracy and discrimination power.

# Experimental results and discussion

## Correlation Matrix Heatmap




The heatmap visually represents relationships between variables, identifies patterns and associations between variables. The correlation matrix shows output variable 'Risk' has strong correlations with variables like 'Q227_Blame_myself', 'Q225_Feel_lonely', and 'Q228_Not_Worth_living', suggesting that these variables are closely related to "Risk".

## Chi-square Test

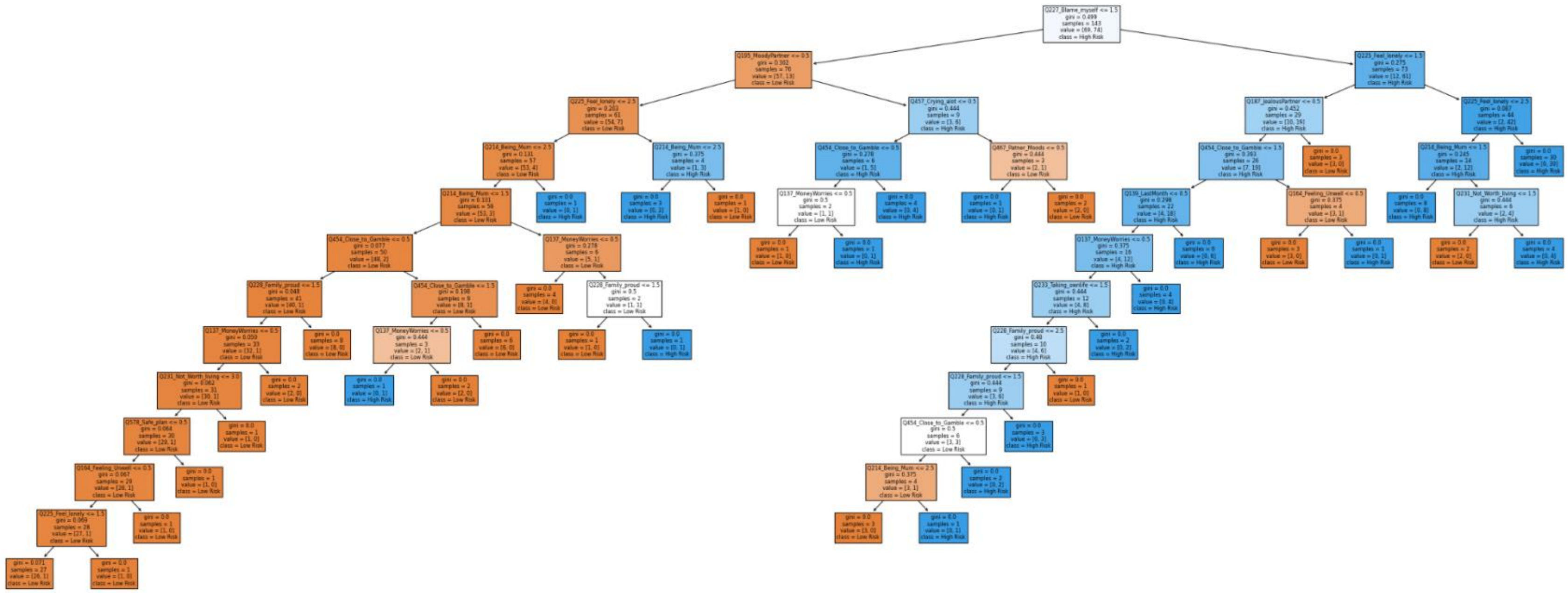
| Column | Chi-square | p-value |
|---|---|---|
| Q214_Being_Mum | 21.633755 | 7.773811e-05 |
| Q225_Feel_lonely | 56.827451 | 1.344702e-11 |
| Q227_Blame_myself | 85.379777 | 1.260079e-17 |
| Q228_Family_proud | 26.371369 | 7.973634e-06 |
| Q231_Not_Worth_living | 42.608036 | 1.247859e-08 |
| Q233_Taking_ownlife | 18.811216 | 8.559804e-04 |
| Q454_Close_to_Gamble | 11.977436 | 2.506876e-03 |
| Q578_Safe_plan | 9.375036 | 2.199604e-03 |
| Q589_SP_Patner | 5.710314 | 1.686553e-02 |
| Q535_SP_Clincian | 11.109256 | 8.589796e-04 |
| Q536_SP_keep_safe | 7.264182 | 7.034331e-03 |
| Q534_SP_Client | 17.407815 | 3.015834e-05 |
| Q588_SP_SafeToGoHome | 7.200887 | 7.286754e-03 |
| Q467_Patner_Moods | 6.310506 | 1.200246e-02 |
| Q137_MoneyWorries | 6.824174 | 8.993212e-03 |
| Q139_Transport | 9.375036 | 2.199604e-03 |
| Q174_Feeling_Alone | 7.568958 | 5.938210e-03 |
| Q164_Feeling_Unwell | 10.318842 | 1.316789e-03 |
| Q457_Crying_alot | 11.846283 | 5.777656e-04 |
| Q196_AngryPartner | 9.177946 | 2.449488e-03 |
| Q187_Cp_Partner | 6.236015 | 1.251778e-02 |
| Q187_JealousPartner | 6.167653 | 1.301071e-02 |
| Q195_MoodyPartner | 10.882299 | 9.708767e-04 |

According to the chi-square test results, some variables that are highly significant are:

- Q227_Blame_myself
- Q225_Feel_lonely
- Q231_Not_Worth_living
- Q228_Family_proud

These variables have low p-values, suggesting a strong association and most significant at predicting “Risk”.

DT



In DT, blue represents the classification of "High Risk," while the orange represents the classification of "Low Risk." The tree branches out based on different features and their corresponding thresholds to make predictions.

The feature "Q227_Blame_myself" is used to split the data at the root node. If its value is below or equal to a threshold, the path goes to the left child node, potentially indicating a higher chance of "High Risk" classification. If the value is above the threshold, the path goes to the right child node, indicating a higher likelihood of "Low Risk" classification.

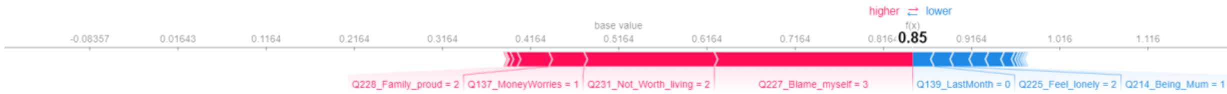
This variable, along with other variables such as "Q225_Feel_lonely," "Q187_JealousPartner," "Q454_Close_to_Gamble," "Q139_Transport," "Q137_MoneyWorries," and "Q228_Family_proud," are potential factors contributing to the classification of "High Risk" in the dataset.

The Gini impurity at the root node is 0.499, indicating the level of impurity or the likelihood of misclassifying a randomly selected sample at this point. A lower Gini impurity suggests a more homogeneous node.

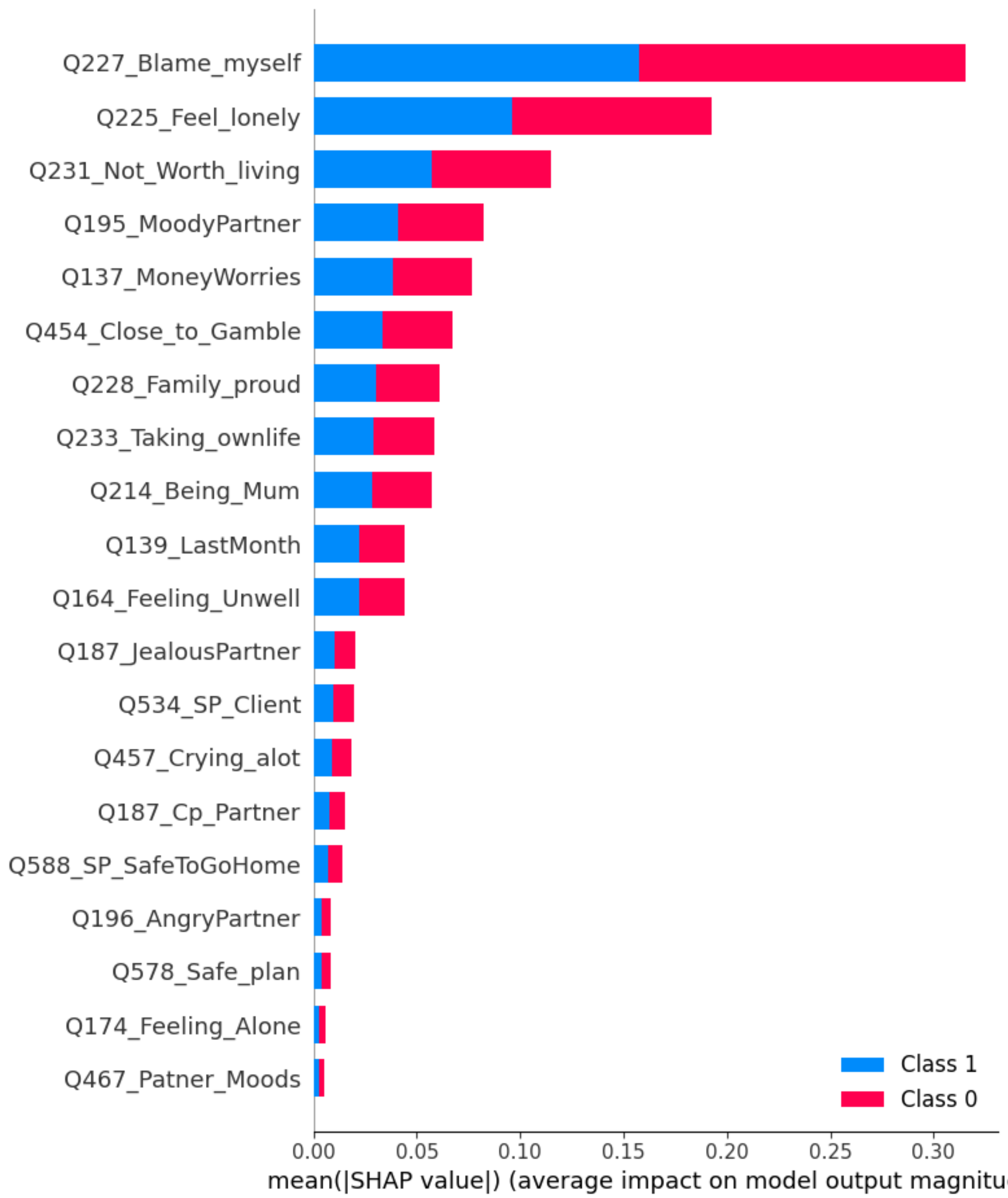
Furthermore, out of the 143 samples that reach this node, approximately 69.74% are classified as "High Risk." This suggests that the variables mentioned earlier may play a significant role in predicting the occurrence of "High Risk" cases.

In summary, the DT suggests that factors such as self-blame, feelings of loneliness, jealousy in relationships, involvement in gambling, transportation issues, financial concerns, and family pride are potential indicators of a higher risk of mental health issues.

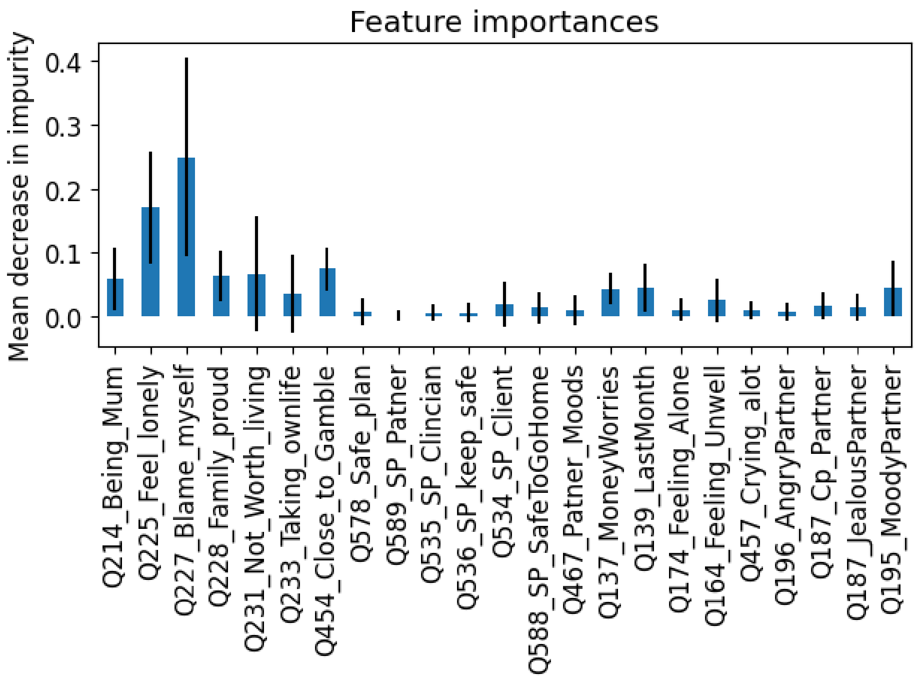
<u>RF interpretation with SHAP</u>




Clinician may select a single case for interpretation. For example, for row 100, model predicted 0.85 (towards “High_risk”), with base value 0.5164. The largest effect is “Q227_Blame_myself”.

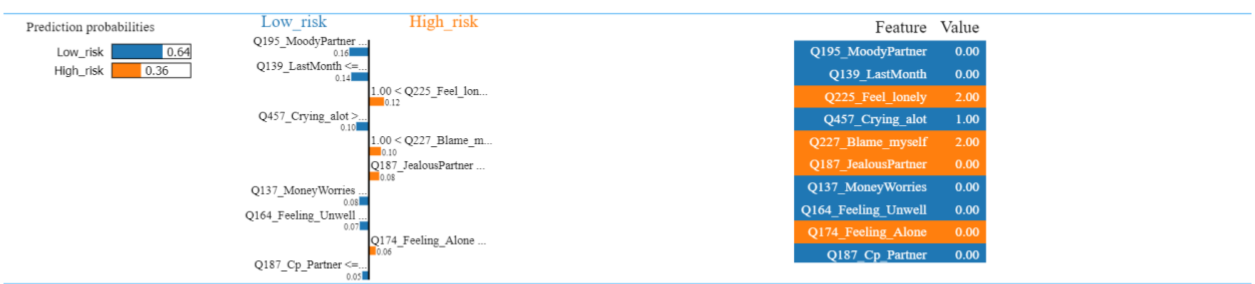


The most important features are shown from top to bottom. “Q227_Blame_myself” has the highest impact on both classes, followed by “Q225_Feel_lonely” and “Q231_Not_Worth_living”.

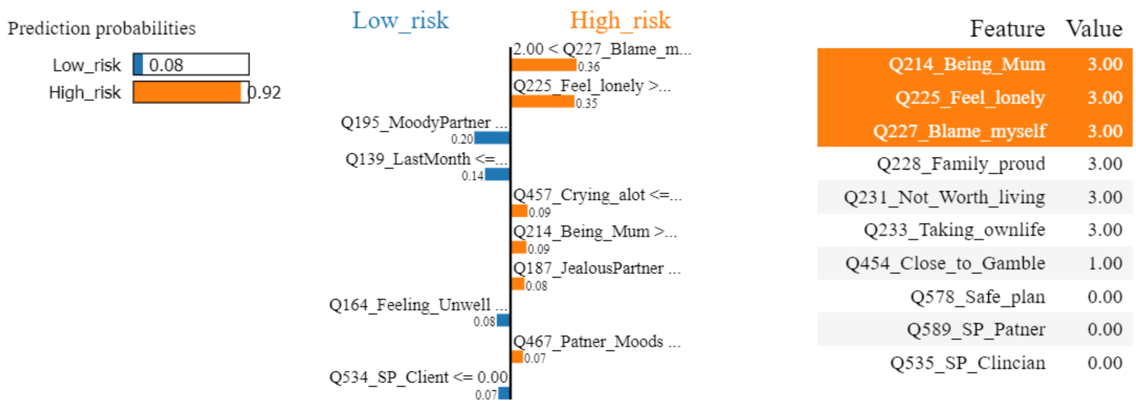



Feature importance (on mean decrease in impurity) shows similar results.

## LR interpretation using LIME




Similarly, clinician by analyse row 5 from X_test, recognising LR model predicts “Low_risk” with 64% confidence, explaining prediction with “Q195_Moody_Partner” and “Q139_LastMonth” as primarily contributing factors. However, these alone does not explain its different classification from the below case.




# Conclusion

In conclusion, models used in our study demonstrated reasonable outcome of predicting “Risk”. The most interpretable model is DT, followed by either RF or LR. DT diagram is plotted and at each step, the clinician may verify how decisions are made. RF is useful in identification of important features, although as explained before, its definition varies. We would recommend DT due to its inherent interpretability and explainability.

There are limitations to our project:

1. Little involvement of using patient-defined outcome measures since caring for infant may contribute to perinatal mental health disorders. Therefore, sample of safety which is reliant on long-term results are difficult as is rarely accumulated.
2. Although the above outcomes demonstrated the most probable risk that could lead to mental health issues, these do not guarantee that there aren’t other factors that could contribute to “Risk” in reality.
3. Several possible overlapping factors due to questions asked and survey methodology is questionable.

In general, more research should be conducted to broaden study of the project, to produce a more reliable outcome with higher accuracy (important for identification to doctors for treatment).

In future, we plan to manually remove unimportant or similar features with little impact on the prediction classification.

# References

[1] C. O’Sullivan, ‘What are Model Agnostic Methods?’, *Medium*, Mar. 15, 2023. Available: https://towardsdatascience.com/what-are-model-agnostic-methods-387b0e8441ef. [Accessed: Jun. 12, 2023]

[2] National Health Service, ‘Perinatal mental health’. Available: https://www.england.nhs.uk/mental-health/perinatal/. [Accessed: Jun. 11, 2023]

[3] M. Muzik and S. Borovska, ‘Perinatal depression: implications for child mental health’, *Ment. Health Fam. Med.*, vol. 7, no. 4, pp. 239–247, Dec. 2010.

[4] R. Ramakrishnan, S. Rao, and J.-R. He, ‘Perinatal health predictors using artificial intelligence: A review’, *Womens Health*, vol. 17, p. 17455065211046132, 2021, doi: 10.1177/17455065211046132

[5] A. Bertini, R. Salas, S. Chabert, L. Sobrevia, and F. Pardo, ‘Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review’, *Front. Bioeng. Biotechnol.*, vol. 9, 2022, doi: 10.3389/fbioe.2021.780389. Available: https://www.frontiersin.org/articles/10.3389/fbioe.2021.780389. [Accessed: Jun. 11, 2023]

[6] L. M. Howard and H. Khalifeh, ‘Perinatal mental health: a review of progress and challenges’, *World Psychiatry*, vol. 19, no. 3, pp. 313–327, 2020, doi: 10.1002/wps.20769

[7] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, and L. Cilar, ‘Interpretability of machine learning-based prediction models in healthcare’, *WIREs Data Min. Knowl. Discov.*, vol. 10, no. 5, p. e1379, 2020, doi: 10.1002/widm.1379

[8] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020.

[9] Y. Zhang, S. Wang, A. Hermann, R. Joly, and J. Pathak, ‘Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women’, *J. Affect. Disord.*, vol. 279, pp. 1–8, 2021, doi: 10.1016/j.jad.2020.09.113

[10] D. Gunning and D. Aha, ‘DARPA’s explainable artificial intelligence (XAI) program’, *AI Mag.*, vol. 40, no. 2, pp. 44–58, 2019, doi: 10.1609/aimag.v40i2.2850

[11] K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, and J. Qadir, ‘Explainable, trustworthy, and ethical machine learning for healthcare: A survey’, *Comput. Biol. Med.*, vol. 149, p. 106043, Oct. 2022, doi: 10.1016/j.compbiomed.2022.106043