Decision Tree Model for Maternal Risk Classification Based on Kartu Skor Poedji Rochjati

Dyah Megawati Surip Solekhah[1*], Annisa Maulida Ningtyas[2]
[1,2]*Department of Health Information and Services, Vocational College, Universitas Gadjah Mada*
*corr-author: dyah.megawati0402@mail.ugm.ac.id

Abstract - Maternal mortality remains a serious public health issue. This can be prevented by taking preventive measures through early identification of pregnancy risks. This study aims to develop a classification model for screening maternal pregnancy risks using the "*Kartu Skor Poedji Rochjati*" (KSPR) as a clinical basis for data labelling. This research applies the Cross-Industry Standard Process for Medical Data Mining (CRISP-MED-DM) methodology to ensure a systematic and clinically relevant modelling process. A dataset containing 998 medical records from the Maternal Health and High-Risk Pregnancy Dataset was used, with rule-based labelling adapted from KSPR to classify maternal risk into three categories: Low-Risk Pregnancy, High-Risk Pregnancy, and Very High-Risk Pregnancy. Experiments were conducted with three Decision Tree models, namely the Baseline Model, Model with SMOTE, and Model with Class Weighting. Based on these experiments, it was found that the Decision Tree algorithm enhanced with the Synthetic Minority Oversampling Technique (SMOTE) to overcome class imbalance was the most optimal model. This model achieved balanced performance across all classes with an accuracy of 0.86 and a weighted average F1-Score of 0.87.

Keywords: Maternal health, decision tree, *Kartu Skor Poedji Rochjati*, SMOTE, class weighting.

I. INTRODUCTION

Pregnancy is a condition that requires special attention because even normal pregnancies carry potential risks that may endanger the lives of both the mother and fetus [1]. The global maternal mortality ratio (MMR) in 2020 was 223 deaths per 100,000 live births [2]. The Indonesia Central Statistics Agency reported that the maternal mortality rate (MMR) reached 189 per 100,000 live births in 2020 [3]. In addition, the Ministry of Health also reported that there were 4,482 maternal deaths throughout Indonesia in 2023. Therefore, efforts are needed to reduce maternal mortality, which is also one of the key targets in the Sustainable Development Goals (SDGs), to lower the global MMR to 70 deaths per 100,000 live births by 2030 [4].

Preventive efforts through comprehensive maternal health services are essential to minimize the risk of maternal death. One preventive step that can be taken is early detection of related risk in pregnant women [5]. In Indonesia, one of the instruments widely used for antenatal screening is the *Kartu Skor Poedji Rochjati* (KSPR), which is a scoring system to assess maternal risk based on 20 criteria related to disease that endanger pregnant women such as maternal age, parity, height, and pre-existing diseases like anemia [6]. This scoring system categorizes risk levels into three categories, Low-Risk Pregnancy (LRP), High-Risk Pregnancy (HRP), and Very High-Risk Pregnancy (VHRP). However, the manual recording process using the KSPR can lead to inefficiency, recording errors, and limitations in data analysis [7]. The manual recording process has the potential to cause recording errors because the process is repetitive and depends on human accuracy, thus allowing for human error [8]. In addition, data recorded on paper requires re-entry into the computer system before it can be processed and analysed, making it less efficient. The limitations of the manual recording process using the KSPR screening can be overcome by utilizing technology. With advances in technology, computational methods have been widely used to improve diagnostic accuracy and assist clinicians in decision-making [9].

One such approach is the Clinical Decision Support System (CDSS), which integrates machine learning to support medical decision-making. The application of machine learning in healthcare is growing to enhance diagnosis and risk prediction [10]. The main principle of machine learning is the development of algorithms that utilize statistical analysis on input data to generate output predictions [11]. Several CDSS models have been developed using various machine learning dan deep learning algorithms such as Support Vector Machine (SVM), Random Forest, and various deep learning method such as Convolutional Neural Network [12,13]. Although Neural Networks (NN) are better for complex, high-dimensional, and nonlinear data [14]. This algorithm also has high accuracy in complex tasks like medical image analysis, but their 'black box' system

poses challenge that needs to be addressed. The inability to trace the internal logic of these models can lead to a lack of trust from clinicians, who require transparent and verifiable reasoning, especially when decisions impact patient safety [15].

Among the various machine learning algorithms, Decision Tree is one of the most suitable choices because it is transparent in its decision-making. This is especially important because patients and clinicians have limited trust in artificial intelligence (AI) algorithms, which are sometimes perceived as "black box", where inputs produce outputs without clear understanding of the internal process [16]. This distrust in AI algorithms, particularly concerning its accuracy and reliability where decisions can determine the life or death of the patient. The hierarchical structure of a decision tree can be easily translated into a series of 'IF-THEN' rules [17]. This property places Decision Trees in the category of inherently interpretable models [18]. A previous study applied the Classification and Regression Tree (CART) algorithm to the Maternal Health Risk dataset from the UCI Machine Learning Repository, which produced a model with an accuracy of 90.57% [19]. Although the accuracy of the research model had quite good results, the study used labels inherent in the dataset that were created based on clinical standars outside Indonesia. The use of the dataset's default labels may not reflect the standars for pregnancy risk classification in Indonesia. Therefore, the application of this model in Indonesia healthcare system may have limitations.

To address these challenges, this study aims to develop a Decision Tree-based classification model specifically designed for maternal risk screening using features that are in line with KSPR, which is widely used in Indonesia for identifying pregnancy risk levels. This study also emphasizes the adaptation of the maternal health dataset to align with KSPR, enabling the development of a model that is more relevant to the clinical context in Indonesia. Furthermore, the Decision Tree structure enables the model to represent clinical indicators into a series of understandable decision rules that support healthcare professionals in identifying pregnancy risks.

II. METHOD

This study uses the Cross Industry Standard Process Medical for Data Mining (CRISP-MED-DM) method, an adaptation of the standard Cross-Industry Standard Process for Data Mining (CRISP-DM) to meet the specific needs of the medical environment. CRISP-MED-DM was created to support the development of data mining-based systems that are more oriented towards specific health issues [20]. One of the main modifications in this framework is the replacement of the 'Business Understanding' phase, which typically focuses on business objectives, with the 'Problem Understanding' phase [20]. This change emphasizes the need for a deep understanding of the clinical context and medical problems to be solved. The first phase, Problem Understanding, focuses on comprehensively analyzing the main clinical problem, which concerns the high maternal mortality rate (MMR) in Indonesia and the important role of early risk detection as a preventive measure.

A. Dataset

This study uses Maternal Health and High-Risk Pregnancy Dataset, obtained from IEEE DataPort [21]. A total of 998 medical records were collected from September 2023 to November 2024 at the Tala Surgical Clinic in Bangladesh. Each record includes various demographic and clinical attributes, such as maternal age, number of previous pregnancies, blood pressure, fetal position, and history of complications. This dataset was selected due to the limited availability of openly accessible maternal health datasets from Indonesia. In addition, patient medical records are restricted by privacy regulations and institutional data governance. Although collected in Bangladesh, this dataset contains clinical variables that are relevant and commonly used in pregnancy risk assessment, including maternal age, blood pressure, anemia status, and obstetric history. These indicators are widely recognized as key determinants of pregnancy complications in various healthcare settings and are frequently used in pregnancy risk screening frameworks. Therefore, this dataset provides a suitable basis for developing and evaluating machine learning models for pregnancy risk classification. Furthermore, preliminary exploratory data analysis was conducted to understand the general picture of the data. Data visualizations such as histograms and box plots were used to understand the distribution, dispersion, and detect outlier values in each variable.

B. Data Preparation

Raw data requires several preprocessing steps to ensure its quality and suitability for making a robust model. This preparation process is essential for handling missing values, data inconsistencies, and processing features that are appropriate for the clinical context. The main stages of data preparation are as follows:

1) Feature Selection: This step involved synchronizing the variable in the dataset with the KSPR

instrument. Variables that did not relate to the KSPR instrument were removed. The variables used in this study were selected based on their relevance to the pregnancy risk indicators contained in the *Kartu Skor Poedji Rochjati* (KSPR) instrument. Variables representing clinically recognized maternal risk factors in KSPR, such as maternal age, gravida status, height, blood pressure, anemia, and fetal position, were retained as features in the modeling process. Conversely, variables that could not be mapped to KSPR risk indicators or were not directly related to the pregnancy risk screening process were eliminated from the dataset. In addition, several variables were also removed because they had unclear definitions or contained sensitive patient identity information. This feature selection process was carried out to ensure that the model developed used clinically relevant predictors. A detailed breakdown of the selection process is shown in Table I.

2) Handling Inconsistent Values: The value "none" in anemia_disease variable was recorded as "negative". The "minimal" and "medium" values were consolidated into a "positive" category to create a binary feature.

3) Data Duplicate Removal: A total of 42 duplicate rows were identified and subsequently removed from the dataset. Duplicate records may lead to the overrepresentation of certain observations, which can affect the learning process and distort model evaluation results. After removing these duplicate entries, the final dataset used for subsequent analysis consisted of 956 records.

4) Feature Engineering: There were several things in these steps. First, the blood_pressure variable, originally a text string, was parsed into two separate numeric columns: systolic pressure and diastolic pressure. Second, the result from the previous step made a new categorical feature, blood_pressure_category to provide a clinically relevant variable for the model. Last, values in the height variable were converted from feet to centimetres to adjust with the measurement standard used in the KSPR framework.

TABLE I
VARIABLE SYNCHRONIZATION

No.	Variable in Dataset	Relevance	Decision	Description
1	Name	No	Deleted	Contains patient privacy data
2	Mother's Age	Yes	Used	Related to factors of young primiparity (<16 years) and old primiparity (>35 years)
3	Gravida	Yes	Used	Associated with multiparity risk factors
4	TiTi Tika	No	Deleted	Variable definition and data values could not be identified
5	Gestational Age	Yes	Used	Associated with prolonged pregnancy risk
6	Mother's Weight	No	Deleted	Not included as a risk indicator in the KSPR
7	Height	Yes	Used	Related to narrow pelvis risk if <145 cm
8	Blood Pressure	Yes	Used	Clinical indicator for preeclampsia
9	Anemia	Yes	Used	Represents maternal health condition that associated with pregnancy complications
10	Jaundice	No	Deleted	Not included as a risk indicator in the KSPR
11	Fetal Position	Yes	Used	Related to breech/transverse position
12	Fetal Movement	No	Deleted	Not included as a risk indicator in the KSPR
13	Fetal Heart Rate	No	Deleted	Not included as a risk indicator in the KSPR
14	Albumin Level	No	Deleted	Not listed as a primary risk factor in the KSPR instrument
15	Glucose Level Test	No	Deleted	Not listed as a primary risk factor in the KSPR instrument
16	VDRL Test Results	Yes	Used	Represents maternal health condition that associated with pregnancy complications
17	HbsAg Test Results	Yes	Used	Represents maternal health condition that associated with pregnancy complications
18	Pregnancy Risk Level	No	Deleted	Removed because the risk level was redefined based on the KSPR scoring adaptation

C. Labelling Data Based on KSPR

A constructed rule-based labelling approach was employed by adapting the scoring system of the *Kartu Skor Poedji Rochjati* (KSPR), a clinical tool used for early detection of high-risk pregnancies in Indonesian healthcare. A cumulative risk score was calculated based on clinical attributes such as maternal age, systolic and diastolic blood pressure, fetal position, and relevant medical history from each patient's medical record. Each feature contributed a risk score incrementally, reflecting its clinical significance according to KSPR guidelines. The total score was used to classify maternal risk levels into three categories, such as LRP, HRP, and VHRP. Thresholds for each category were determined in accordance with the official KSPR scoring rubric. This labelling strategy ensured that the target classes used in the model were clinically grounded and aligned with established maternal health risk assessment practices.

The target variable or risk level label was constructed using a multi-stage labelling approach. The process began by adapting the *Kartu Skor Poedji Rochjati* (KSPR) to the available dataset. The results of the variable synchronization process indicate that the dataset contains only nine of the more than twenty risk factors included in the KSPR instrument. This gap between the comprehensive clinical tool and the available features introduces a fundamental source of model uncertainty, as potential risks outside these nine features could not be directly observed. To account for this inherent uncertainty, as well as to simulate the natural variability in real-world clinical judgment, a two-stage scoring process was designed. The entire process, illustrated in Figure 1, was executed with a fixed random seed (seed=42) to ensure full reproducibility.

The first stage involved calculating a deterministic baseline score (S_{base}) as shown in (1), for each patient record, a base score of 2 was initialized. This score was then incrementally updated based on the nine available clinical risk factors (e.g., maternal age, blood pressure). Each satisfied condition contributed a score of +4 or +8, according to its severity as defined in the KSPR guidelines. This baseline represents the risk assessment based purely on the limited data available.

$$S_{base} = 2 + \sum_{i=1}^{9} I(f_i) \cdot P(f_i) \quad (1)$$

where:

- $I(f_i)$: Is an indicator function that equals 1 if the patient has risk factor f_i and 0 otherwise
- $P(f_i)$: Is the point value (+4 or +8) assigned to risk factor f_i

The second stage introduced a stochastic simulation of uncertainty (ϵ). To computationally represent the net effect of the unobserved KSPR factors and the variability of clinical judgment, a small, random perturbation was added to the baseline score as shown in (2). This was implemented as a discrete random variable, ϵ, with the following probability distribution:

$$\epsilon = \begin{cases} +1 & with\ probability\ 0.15 \\ 0 & with\ probability\ 0.70 \\ -1 & with\ probability\ 0.15 \end{cases} \quad (2)$$

The final adjusted score (S_{total}) for each patient was calculated by combining baseline score and the stochastic component, ensuring a minimum score of 2 as shown in (3).

$$S_{total} = max(2, S_{total} + \epsilon) \quad (3)$$

This technique serves as a proxy to model the "fuzziness" and uncertainty that exist in a real clinical setting but cannot be directly calculated from the dataset

Finally, the adjusted cumulative score was mapped to one of three risk categories using predefined thresholds: LRP, HRP, and VHRP. This methodology yields a target variable that is not only grounded in a clinical framework but also acknowledges and attempts to model the inherent limitations and uncertainties of the research context. The final mapping represented in (4).

$$\text{Risk Category} = \begin{cases} \text{Low Risk Pregnancy (LRP);} & \text{if } S_{total} < T_{HRP} \\ \text{High Risk Pregnancy (HRP);} & \text{if } T_{LRP} \le S_{total} < T_{VHRP} \\ \text{Very High Risk Pregnancy (VHRP);} & \text{if } S_{total} \ge T_{VHRP} \end{cases} \quad (4)$$

D. Modelling

To construct the predictive model for pregnancy risk classification, a Decision Tree classifier was employed as the primary algorithm. This choice was motivated by its high interpretability, as it generates rule-based decision structures that closely resemble clinical reasoning. The model development process was designed as a comparative experiment to evaluate different strategies for handling the severe class imbalance observed in the dataset.

1) Baseline Model: This model is a standard Decision Tree classifier, and it was trained on the imbalanced training data. This model serves as a benchmark to measure the impact of imbalance-handling technique.

2) Model with SMOTE: This model was integrated with Synthetic Minority Oversampling technique (SMOTE) to the workflow. SMOTE was applied to the training data within scikit-learn pipeline to prevent data leakage and ensuring that synthetic samples were generated only during the training phase of each cross-validation fold.

3) Model with Class Weighting: This is an alternative approach that is implemented by adjusting the class_weight hyperparameter of the Decision Tree Classifier to 'balanced'. This method modifies the algorithm's learning process by assigning higher penalties to misclassifications of the minority class (VHRP), thereby forcing the model to pay more attention to it without altering the dataset itself.

For all three experimental setups, the dataset was partitioned into a training set (80%) and a testing set (20%) using stratified sampling. To ensure a fair comparison, each model underwent the same hyperparameter optimization process using a grid search with 10-fold stratified cross-validation. The key hyperparameters tuned included max_depth and min_samples_leaf, with the objective of finding the optimal balance between complexity and predictive performance for each approach.

E. Evaluation

The Decision Tree model evaluation process uses several classification metrics, including accuracy, precision, recall, and F1 score. In this multi-class context, accuracy provides an overall measure of correct classification, while precision measures the reliability of positive predictions and recall measures the model's ability to identify all true positive cases. Due to significant class imbalance in the dataset, particularly in the Very High-Risk Pregnancy class, additional evaluation metrics are used, namely macro average and weighted average schemes. This approach provides a more appropriate assessment than single classification metrics alone. Macro average calculates the unweighted average of metrics across all classes, treating each class equally regardless of its size. On the other hand, weighted average considers class prevalence by weighting each class's score based on its frequency in the dataset, providing a view of performance relative to the overall class distribution.

III. RESULT AND DISCUSSION

A. Dataset Characteristic

This study utilized the "Maternal Health and High-Risk Pregnancy" dataset, from the IEEE DataPort repository. The dataset comprises 998 records of pregnant women in Bangladesh. During the data preprocessing stage, a total of 42 duplicate records were identified and removed to eliminate redundant observations that could potentially bias the model training process and evaluation results. After this cleaning process, the final dataset used for subsequent analysis consisted of 956 records. From the initial set of available variables, nine features were selected to align with the clinical risk factors outlined in the *Kartu Skor Poedji Rochjati* (KSPR) framework, a standard screening tool in Indonesia. This selection ensures the clinical relevance of the model's input variables. The selected features and their descriptions are presented in Table II.

An exploratory data analysis was conducted to understand the dataset's underlying structure. The descriptive statistics for the numerical features are summarized in Table III, showing a maternal age range from 15 to 50 years, with an average age of approximately 23.

The primary focus of the analysis was the distribution of the target variable, risk_label, which was generated using the KSPR labeling methodology described in Section 3.3. The distribution of the three risk classes is detailed in Table III and visualized in Fig. 2.

The analysis of the labeled data revealed a significant class imbalance within the dataset. The HRP class constitutes the majority with 50.4% of the records (482 out of 956 patients), followed closely by the LRP class at 45.1% (431 out of 956 patients). Critically, the VHRP class is severely underrepresented, accounting for only 4.5% of the data (43 out of 956 patients). This imbalance poses a critical challenge for model development, as a standard classifier trained on this data would likely become biased towards the majority classes and perform poorly in identifying the minority high-risk cases, which are the most crucial to detect clinically. This finding directly justifies the use of a resampling technique, such as SMOTE, in the subsequent model development phase to mitigate this bias.

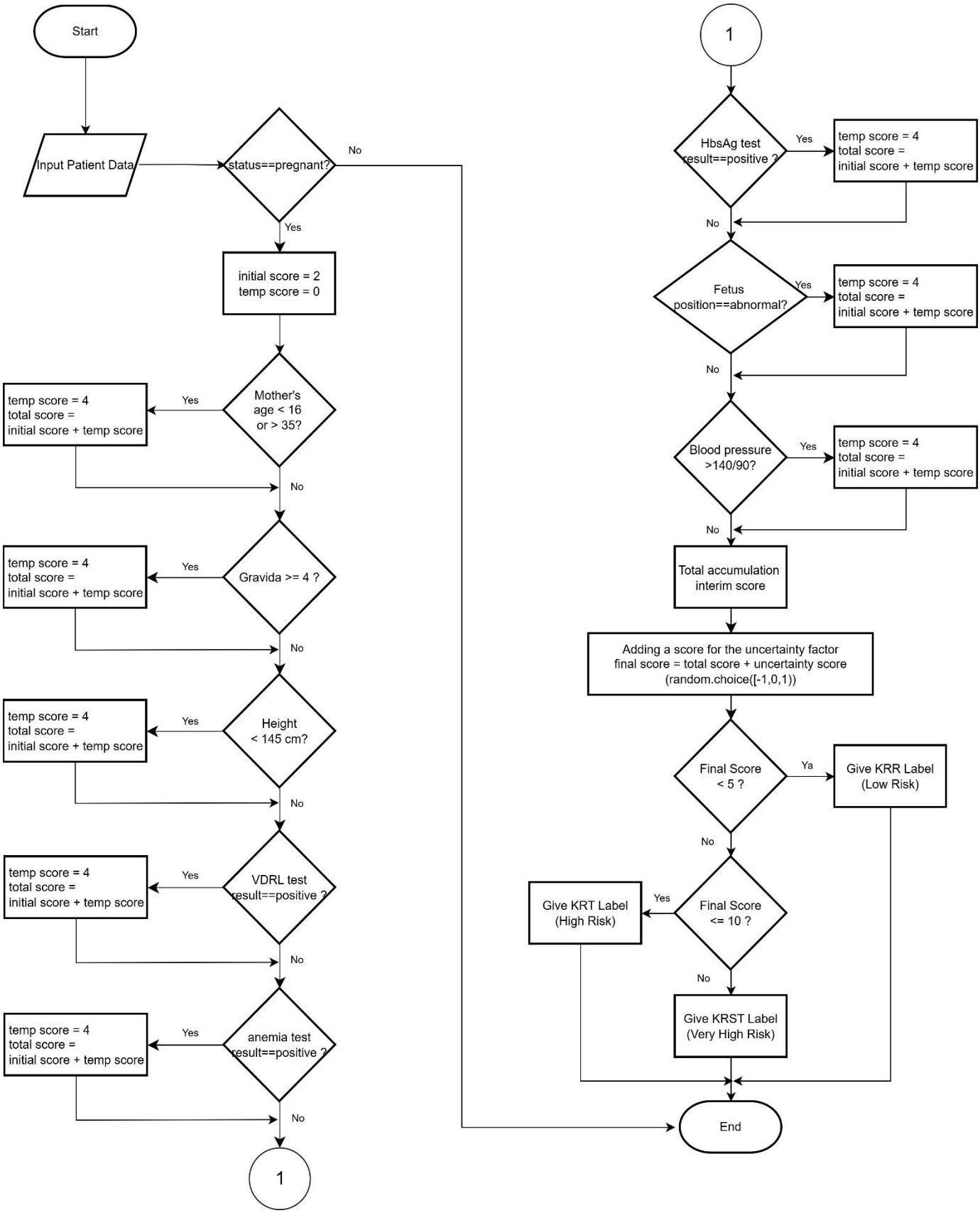

Fig. 1 Data labelling flowchart

TABLE II
DESCRIPTION OF SELECTED FEATURES

No	Feature Name	Data Type	Description
1	mother's_age	Numerical	Maternal age in years
2	gravida	Numerical	Number of previous pregnancies
3	gestational_age	Numerical	Gestational age in weeks
4	mother's_age height	Numerical	Maternal height in centimeters
5	anemia_disease	Categorical	Presence of anemia (Positive/Negative)
6	fetal_position	Categorical	Fetal presentation (e.g., Normal, Abnormal)
7	VDRL_test_results	Categorical	VDRL test result (Positive/Negative)
8	HbsAg_test_results	Categorical	HBsAg test result (Positive/Negative)
9	blood_presure	Categorical	Blood pressure

TABLE III
DESCRIPTIVE STATISTIC FOR NUMERICAL FEATURES

Statistic	mother's_age	gravida	gestational_age	height	systolic_pressure	diastolic_pressure
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000
mean	22.995816	1.188285	28.883891	158.643724	102.913180	64.623431
std	3.617292	0.441450	5.018511	4.382737	9.681406	8.055075
min	17.000000	1.000000	20.000000	142.240000	80.000000	55.000000
25%	20.000000	1.000000	25.000000	154.940000	100.000000	60.000000
50%	22.000000	1.000000	30.000000	157.480000	100.000000	60.000000
75%	25.000000	1.000000	32.000000	162.560000	110.000000	70.000000

B. Model Performance

The comparative experiment was conducted to identify the most effective strategy for handling the dataset's class imbalance. The performance of the three models, namely a baseline Decision Tree, a model with SMOTE, and a model with class_weight was evaluated using both class-specific and aggregated metrics. A detailed comparison of the F1-Scores for each risk category reveals distinct performance trade-offs among the models (Table IV). The baseline Decision Tree model, despite being trained on the original imbalanced data, achieved the highest F1-Score for the critical VHRP class at 0.74. In contrast, the model incorporating SMOTE showed better performance on the majority classes, attaining the highest F1-Scores for both LRP and HRP at 0.90 and 0.86, respectively. The model utilizing the class_weight parameter also performed well on the LRP class (0.90), but its ability to identify the minority class was significantly diminished, resulting in the lowest F1-Score for VHRP (0.51).

When considering the aggregated performance metrics, the model with SMOTE emerged as the most robust overall (Table V). It achieved the highest accuracy (0.86) and the highest weighted average F1-Score (0.87), indicating the best-balanced performance across the imbalanced class distribution. Interestingly, the baseline model recorded the highest macro average score (0.80), which reflects strong average performance without accounting for class imbalance and aligns with its superior handling of the VHRP class. Meanwhile, the model with class_weight obtained the lowest macro average score of 0.75, further highlighting its weakness in managing the underrepresented VHRP category.

The performance of the final Decision Tree model, trained using the SMOTE-enhanced dataset, was evaluated in two stages. First, the model performance was assessed using 10-fold cross-validation on the training dataset. Second, the model's predictive ability was evaluated on an unseen test set obtained from the initial train–test split. In this study, the dataset was divided into training and testing subsets using an 80:20 split, resulting in 764 records used for training and 192 records reserved for testing. The model demonstrated stable performance during the 10-fold cross-validation process. It achieved an average F1-Macro score of 0.755, with scores across the folds ranging from a minimum of 0.613 to a maximum of 0.843, as detailed in Table VI. This relatively narrow range indicates that the model generalizes well across different subsets of the training data and is not overly sensitive to data partitioning, suggesting good stability.

The model's final predictive capability was assessed on the unseen test set, comprising 192 patient records. The overall performance metrics are summarized in Table VII. The model achieved an overall accuracy of 86%, a macro-averaged F1-score of 0.79, and a weighted-averaged F1-score of 0.87.

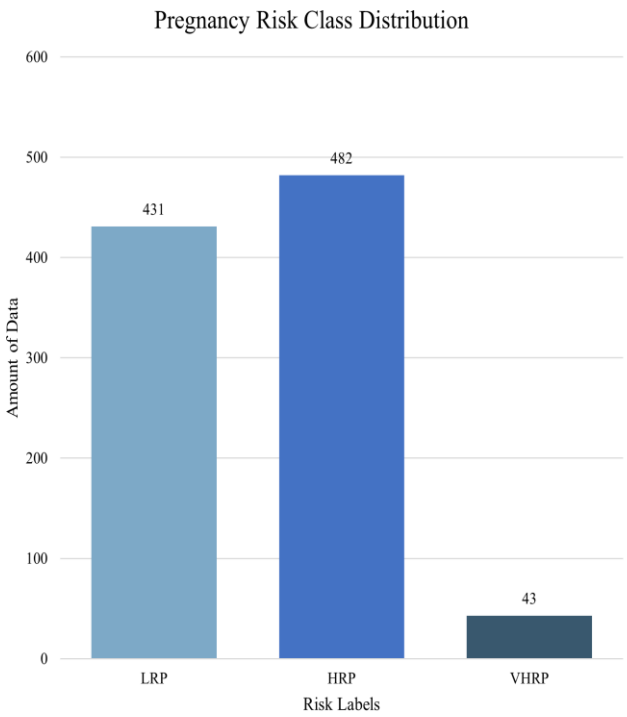

Fig. 2 Distribution of maternal risk classes

TABLE IV
COMPARISON OF F1-SCORE PER CLASS IN THREE MODELS

No.	Model	F1-Score LRP	F1-Score HRP	F1-Score VHRP
1	Baseline Model	0,85	0,83	0,74
2	Model with SMOTE	0,90	0,86	0,60
3	Model with Class Weighting	0,90	0,83	0,51

TABLE V
COMPARISON OF MODEL AGGREGATE METRICS

No.	Model	Accuracy	Macro Average	Weighted Average
1	Baseline Model	0,83	0,80	0,83
2	Model with SMOTE	0,86	0,79	0,87
3	Model with Class Weighting	0,83	0,75	0,85

TABLE VI
CROSS VALIDATION TEST RESULT

No.	Evaluation Metrics	Score
1	Average F1-Macro Score	0.755
2	Lowest F1-Macro Score per Fold	0.613
3	Highest F1-Macro Score per Fold	0.843

TABLE VII
CLASSIFICATION REPORT PER CLASS IN TEST SET

No.	Class	Precision	Recall	F1-Score	Support
1	LRP	1,00	0,83	0,90	86
2	HRP	0,85	0,88	0,86	97
3	VHRP	0,43	1,00	0,60	9

TABLE VIII
OVERALL RESULT

No.	Class	Precision	Recall	F1-Score	Support
1	Accuracy	-	-	0,86	192
2	Macro Avg	0,76	0,90	0,79	192
3	Weighted Avg	0,90	0,86	0,87	192

The most critical finding from the evaluation is the model's performance on the VHRP class. The model achieved a recall of 1.00 for this class, meaning that the model can identifying all 9 actual VHRP cases in the test set. This result is relevant because it indicates the model did not miss any of the most critical high-risk patients (zero false negatives).

However, this high sensitivity came with a trade-off in precision, which was 0.43 for the VHRP class. The confusion matrix reveals that this lower precision was caused by 12 HRP cases being misclassified as VHRP (Fig. 3). In a clinical screening context, this type of error (a false positive for a higher risk category) is more acceptable than a false negative. This outcome is advantageous because it ensures that potentially deteriorating patients are flagged for closer monitoring, aligning with the precautionary principle of providing early warnings, even at the risk of some false alarms. The model also demonstrated strong performance for the other classes. For the LRP class, the model achieved a precision of 1.00, meaning every patient predicted as low risk was indeed low risk. For the HRP class, the model maintained a balanced performance with an F1-score of 0.86.

Although the model achieved perfect recall for the VHRP class, this result should be interpreted with caution. The number of VHRP cases in the test set was relatively small (n = 9), which may influence the evaluation results. In addition, the use of SMOTE during training may encourage the model to classify borderline cases into the minority class. This pattern may also indicate a potential tendency toward overfitting to the minority class, where the model becomes highly sensitive to detecting VHRP cases but may reduce precision.

Furthermore, this pattern may suggest a potential tendency toward overfitting the minority class. When synthetic samples are generated through SMOTE, the

model may learn patterns that are strongly associated with the minority class in the training data. As a result, the model may predict more cases as VHRP in the test set, even when some of them belong to the HRP category. Therefore, while the model demonstrates strong sensitivity for identifying critical cases, further validation with larger and more balanced datasets is necessary to ensure the model's generalizability.

These results align with the broader literature demonstrating machine learning's efficacy in health risk prediction, yet our study offers a distinct contribution through its enhanced clinical relevance for the Indonesian context. While other models, such as that by [19], may report higher overall accuracy on generic datasets, our model's specific adaptation of the KSPR framework provides a foundation for local implementation. Furthermore, by employing a Decision Tree, this research directly addresses the call from [22] for greater interpretability in medical AI, filling a critical gap by delivering a "white-box" model whose logic is transparent to healthcare professionals. The practical implications of this are significant; the model can be integrated into a Clinical Decision Support System (CDSS) to standardize screening, accelerate risk identification, and serve as an effective early warning system. The development of a web-based prototype has already confirmed the technical feasibility of deploying such a tool.

Despite these promising implications, it is important to acknowledge the study's limitations. The dataset was sourced from a single institution in Bangladesh and may not fully represent the demographic and clinical characteristics of the target population in Indonesia. Methodologically, the model is an adaptation based on only nine available KSPR features, introducing a degree of model uncertainty. The use of stochastic noise, while justified as a simulation of clinical variability, remains a proxy rather than a direct measurement. Consequently, the model's performance requires further validation on larger, multi-center local datasets before its generalizability can be fully established.

IV. CONCLUSION

Based on a comparison between the three Decision Tree models developed (basic model, model with SMOTE, and model with Class Weight parameters), the model that integrates SMOTE shows optimal performance. This model achieved a weighted F1 score of 0.87 and a high recall rate in detecting the "Very High-Risk Pregnancy" (VHRP) class. Based on these results, further research should prioritize external validation of this model using Indonesian datasets to ensure its

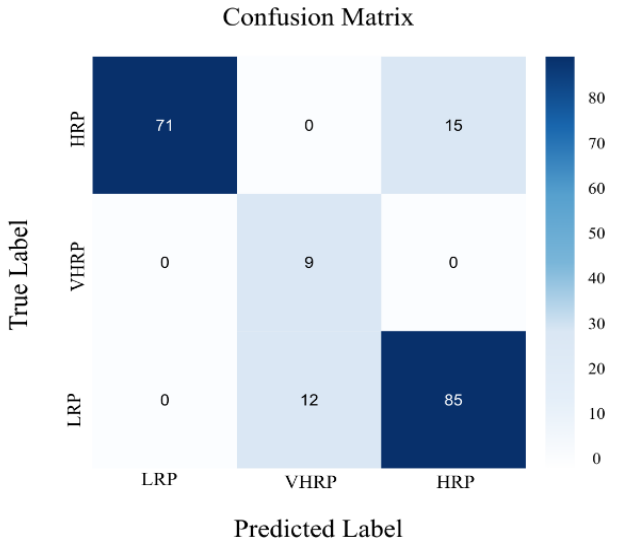

Fig. 3 Confusion matrix

feasibility in the target clinical environment. Along with the advancement of medical record digitization, further research can focus on developing models with a more complete set of KSPR features to improve accuracy. Additionally, subsequent research could explore alternative modeling approaches that maintain a balance between predictive performance and model interpretability. Ultimately, prospective studies designed to evaluate the real-time impact of this tool on clinical decision-making and patient outcomes would be invaluable.

REFERENCES

[1] M. A. Ratnaningtyas and F. Indrawati, "Karakteristik Ibu Hamil dengan Kejadian Kehamilan Risiko Tinggi," *Higeia Journal of Public Health Research and Development*, vol. 3, pp. 334–344, 2023, doi: 10.15294/higeia/v7i3/64147.

[2] World Health Organizations, "Maternal mortality ratio (per 100 000 live births)." Accessed: Jan. 29, 2025. [Online]. Available: https://data.who.int/indicators/i/C071DCB/AC597B1

[3] Badan Sensus Penduduk, *Hasil Sensus Penduduk*. Jakarta: Badan Pusat Statistik, 2020.

[4] United Nations, "The 17 Goals Sustainable Development Goals." Accessed: Jan. 30, 2025. [Online]. Available: https://sdgs.un.org/goals

[5] G. M. Damaraji, "Pendeteksian Jenis Risiko Kehamilan pada Ibu Hamil Menggunakan Algoritma LSTM," 2022.

[6] P. Rochjati, *Skrining Antenatal pada Ibu Hamil*, 2nd ed. Surabaya: Airlangga University Press, 2023.

[7] A. Alkatiri, R. T. N. Handayani, O. Rosa, M. A. Bahruna, and D. P. Arum, "Optimalisasi Pelayanan Posyandu RW 4 Klurak, Candi Melalui Implementasi Sistem Informasi

Aplikasi Web Sikuat Sidoarjo," *KARYA: Jurnal Pengabdian Kepada Masyarakat*, vol. 4, no. 2, pp. 368–373, Aug. 2024, [Online]. Available: https://jurnalfkip.samawa-university.ac.id/KARYA_JPM/article/view/796

[8] M. M. Preethi, A. Bhoomadevi, and A. Amutha, "Electronic Medical Records (EMR) over manual documentation of in-patient records: a scientific insight," 2021.

[9] N. Yudistira, "Peran Big Data dan Deep Learning untuk Menyelesaikan Permasalahan Secara Komprehensif," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 11, no. 2, p. 78, Dec. 2021, doi: 10.36448/expert.v11i2.2063.

[10] H. Habehh and S. Gohel, "Machine Learning in Healthcare," *Curr. Genomics*, vol. 22, no. 4, pp. 291–300, Dec. 2021, doi: 10.2174/1389202922666210705124359.

[11] M. K. Biddinika, A. Masitha, H. Herman, and V. A. N. Fatimah, "Machine Learning Techniques for Heart Disease Prediction Using a Multi-Algorithm Approach," *JUITA: Jurnal Informatika*, vol. 12, no. 2, p. 149, Nov. 2024, doi: 10.30595/juita.v12i2.24153.

[12] A. Yahyaoui, A. Jamil, J. Rasheed, and M. Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, IEEE, Nov. 2019, pp. 1–4. doi: 10.1109/UBMYK48245.2019.8965556.

[13] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, Aug. 2021, doi: 10.1007/s10044-021-00984-y.

[14] H. Taherdoost, "Deep Learning and Neural Networks: Decision-Making Implications," *Symmetry (Basel).*, vol. 15, no. 9, p. 1723, Sep. 2023, doi: 10.3390/sym15091723.

[15] D. Muhammad and M. Bendechache, "Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis," *Comput. Struct. Biotechnol. J.*, vol. 24, pp. 542–560, Dec. 2024, doi: 10.1016/j.csbj.2024.08.005.

[16] T. Hulsen, "Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare," *AI*, vol. 4, no. 3, pp. 652–666, Aug. 2023, doi: 10.3390/ai4030034.

[17] D. J. N. Wong *et al.*, "Developing and Validating Subjective and Objective Risk-Assessment Measures for Predicting Mortality After Major Surgery: An International Prospective Cohort Study," *PLoS Med.*, vol. 17, no. 10, p. e1003253, Oct. 2020, doi: 10.1371/journal.pmed.1003253.

[18] D. Bertsimas and G. A. Margonis, "Explainable vs. interpretable artificial intelligence frameworks in oncology," *Transl. Cancer Res.*, vol. 12, no. 2, pp. 217–220, Feb. 2023, doi: 10.21037/tcr-22-2427.

[19] Z. Farhani, "Sistem Klasifikasi Risiko Kehamilan dengan Algoritma CART (Classification and Regression Tree)," Universitas Gadjah Mada, 2024.

[20] O. Niaksu, "CRISP Data Mining Methodology Extension for Medical Domain. ," 2015.

[21] A. Chayan, "Maternal Health and High-Pregnancy Risk Dataset," 2024, *IEEE DataPort*. doi: https://dx.doi.org/10.21227/ddfa-mf77.

[22] H. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Med. Phys.*, vol. 47, no. 5, May 2020, doi: 10.1002/mp.13764.