EPBI
Journal of Economic Perspectives and Business Innovation
ISSN 3109-2667

Cost Efficiency Management of Early Heart Disease Detection in the Era of Digital Health Service Transformation, A Data Mining Classification Model Analysis Using Naive Bayes and Decision Tree C4.5

Khairunnazri[1*], Yudi Sutaryana[2]
[1,2]Program Studi Sistem Informasi, STMIK Syaikh Zainuddin Nahdlatul Wathan
e-mail: m4sterlenk@gmail.com[1*] yudhi1071@gmail.com[2]

Articel Information

Keywords: Data Mining, Naive Bayes, Decision Tree C4.5, Heart Disease, Cost Efficiency, Digital Health Transformation

Abstract

Cardiovascular disease remains the leading cause of death worldwide and imposes the largest catastrophic financial burden on Indonesia's national health insurance program, yet conventional diagnostic pathways remain costly and time consuming. This study evaluates Naive Bayes and Decision Tree C4.5 for early heart disease detection and quantifies the cost and time efficiency achieved through a data mining based approach compared with conventional examination. Using 300 anonymized patient records from multiple types of Indonesian health facilities, thirteen clinical attributes were used to train and test both classifiers on an identical eighty to twenty stratified split. Naive Bayes achieved an accuracy of 63.33 percent and an F1 score of 62.07 percent, outperforming Decision Tree C4.5, which achieved an accuracy of 55.00 percent and an F1 score of 50.91 percent, although a McNemar test indicated that this difference was not statistically significant. In contrast, the data mining based pathway produced a mean cost efficiency of 84.82 percent per patient and reduced diagnostic turnaround from an average of 8.26 days to approximately 15 minutes, with efficiency remaining consistently high, between 83.28 and 85.86 percent, across all facility types examined. These findings suggest that the primary value of data mining based early detection lies not in the superiority of one classifier over another but in the substantial and broadly distributed reduction in cost and time it offers, supporting its relevance for Indonesia's digital health transformation, particularly at primary care facilities with limited resources.

INTRODUCTION

Cardiovascular disease remains the leading cause of death worldwide, taking an estimated 19.8 million lives in 2022 and accounting for roughly a third of all global deaths, with more than four in five of those deaths resulting from heart attacks and strokes (World Health Organization, 2025). The most recent Global Burden of Disease analysis further shows that cardiovascular mortality has risen sharply over the past three decades, climbing from 13.1 million deaths in 1990 to 19.2 million in 2023, alongside a more than twofold increase in prevalent cases driven largely by population growth and an aging global population (Lindstrom et al., 2022). These figures place cardiovascular disease not only as a pressing clinical concern but also as a growing structural burden on health systems everywhere, particularly in low and middle income countries where access to timely diagnostic services is often limited.

Indonesia illustrates this burden concretely. National health insurance data reported by BPJS Kesehatan show that heart disease was the single most expensive catastrophic illness financed by the program throughout 2024, with more than 22.5 million cases absorbing approximately Rp19.25 trillion, a figure that surpassed the combined cost of cancer, stroke, and kidney failure treatment for the same year (Kompas.com, 2025. When lost productivity among patients of working age is added to direct treatment costs, the total economic burden of heart disease in Indonesia has been estimated at more than Rp67 trillion in a single year (Harian Kompas, 2024). Such figures underscore that the cost of cardiovascular disease is not confined to the clinical setting alone but extends into national fiscal sustainability, making the search for more efficient diagnostic pathways an economic priority as much as a medical one.

A central driver of this cost burden is the nature of conventional cardiovascular diagnosis itself, which typically depends on a sequence of clinical examinations, laboratory tests, and imaging procedures that are time consuming, resource intensive, and in many cases only accessible at higher tier health facilities. This pattern is well documented in the broader literature on cardiovascular care in low and middle income countries, where systematic evidence shows that diagnostic tests such as cholesterol screening are often prohibitively expensive for patients and function as a financial gateway that determines whether treatment can be accessed at all (Mulure et al., 2024). A systematic review of health economic evidence on early cardiovascular detection found that the majority of documented early detection strategies were cost effective compared with standard or delayed detection pathways, yet also noted that the real world economic value of any given strategy depends heavily on the target country and local healthcare context (Oude Wolcherink et al., 2023). This finding is particularly relevant for a country such as Indonesia, where health facilities range from urban tertiary hospitals to rural community health centers with markedly different resource availability, suggesting that any proposed efficiency gain must be examined empirically rather than assumed from evidence generated in higher income settings.

Advances in digital technology have opened a promising avenue for addressing this diagnostic bottleneck through the broader movement toward digital transformation in healthcare. The accumulation of large volumes of structured clinical data, ranging from patient vital signs to laboratory results, has created fertile ground for data mining techniques to support clinical decision making, with applications spanning predictive analytics, population health management, and the electronic transformation of raw health data into actionable clinical knowledge (Yadav et al., 2023). This trend is corroborated by a broader survey of healthcare data mining applications, which found that classification and prediction techniques were among the most frequently applied data mining tasks in the medical domain, spanning diagnosis support, disease risk prediction, and hospital resource management (Shirzad et al., 2021). Within this broader shift, classification algorithms have emerged as one of the most extensively applied data mining tools in clinical informatics, since they allow historical patient records to be transformed into predictive models capable of flagging disease risk before conventional symptoms fully manifest, thereby shortening the diagnostic pathway that would otherwise rely solely on sequential physical and laboratory examination.

Among the various classification techniques applied to cardiovascular diagnosis, Naive Bayes and decision tree based algorithms such as C4.5 have received sustained attention in the data mining literature because of their comparatively low computational cost and their capacity to produce interpretable diagnostic rules from routinely collected clinical attributes. Latha and Jeeva (2019) demonstrated that ensemble variants built

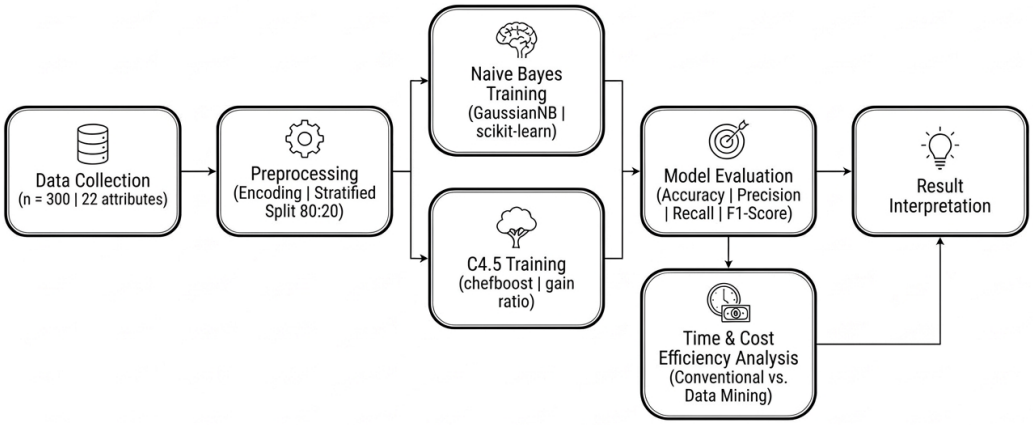
upon these baseline classifiers could meaningfully improve heart disease prediction accuracy on the widely used Cleveland heart disease dataset, while a more recent comparative study evaluating J48, Random Forest, and Naive Bayes on a hospital derived Libyan dataset found that classification accuracy varied considerably across algorithms and improved further once irrelevant attributes were removed through feature selection (Alariyibi et al., 2023). In a similarly structured comparison using the Cleveland dataset, Ananey-Obiri and Sarku (2020) reported that a Gaussian Naive Bayes classifier achieved an accuracy of 82.75 percent, comparable to a logistic regression baseline and noticeably higher than the decision tree model evaluated in the same study, illustrating that the relative ranking between Naive Bayes and tree based classifiers is not fixed but tends to shift depending on the specific dataset and attribute set used. Collectively, these studies confirm that Naive Bayes and C4.5 style decision trees remain relevant, well benchmarked options for heart disease classification even as more complex ensemble and deep learning methods have entered the field.

Despite this substantial body of work, most existing studies on data mining for heart disease diagnosis evaluate classifiers almost exclusively through statistical performance indicators such as accuracy, sensitivity, and specificity, with comparatively little attention paid to translating those performance gains into concrete cost and time savings for the health system. Ahmed et al. (2025) argued that artificial intelligence driven early detection holds real potential to reduce the financial burden of cardiovascular care, yet acknowledged that empirical evidence quantifying this cost impact in operational terms remains limited, particularly outside high income healthcare systems. This limitation carries particular weight for Indonesia, where a nationwide assessment of information and communication technology readiness across primary health care facilities found that digital infrastructure and system maturity at the Puskesmas level generally sit between a basic and a good level, underscoring the need for diagnostic tools that are not only accurate but also lightweight enough to be feasible within existing infrastructure constraints (Aisyah et al., 2024). This gap is even more pronounced in the Indonesian context, where, to the knowledge of the author, no prior study has directly compared the monetary and time cost of a data mining based early detection approach against the cost structure of conventional heart disease examination using locally sourced clinical and financial data.

The present study addresses this gap by developing and comparing Naive Bayes and Decision Tree C4.5 classification models for the early detection of heart disease using a dataset of clinical records drawn from multiple types of Indonesian health facilities, and by explicitly quantifying the cost and time efficiency achieved through the data mining based approach relative to conventional examination procedures. Specifically, this research aims to, first, build and evaluate the classification performance of Naive Bayes and Decision Tree C4.5 models in detecting heart disease based on standard clinical attributes, and second, to analyze the extent of cost and time efficiency generated by the data mining based detection approach compared with conventional diagnostic pathways. By linking algorithmic performance directly to measurable economic outcomes, this study is intended to contribute both to the information systems literature on healthcare data mining and to the practical evidence base needed to support digital transformation policy in the Indonesian health sector. The remainder of this article is organized following the Methods, Results, and Discussion structure, beginning with a description of the dataset, preprocessing steps, and modelling procedure used to generate the classification and cost efficiency results reported in the subsequent sections.

METHODS

This study adopts a quantitative comparative design combined with a data mining oriented workflow following the Cross Industry Standard Process for Data Mining, a process model that structures a data mining project into business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Wirth & Hipp, 2000). Within this workflow, two supervised classification algorithms, Naive Bayes and Decision Tree C4.5, are built on the same dataset and evaluated under identical validation conditions, after which their classification outcomes are examined alongside the recorded cost and time figures in the dataset to assess the economic efficiency of a data mining based early detection approach relative to conventional heart disease examination. Figure 1 illustrates the overall research flow adopted in this study, beginning with data collection and preprocessing, followed by parallel model training for Naive Bayes and Decision Tree C4.5, and concluding with model evaluation and cost and time efficiency analysis.



Figur 1 Research Flow for Heart Disease Early Detection Using Naive Bayes and Decision Tree C4.5

As shown in Figure 1, the two classification algorithms are trained in parallel on an identical data partition, which ensures that any performance difference observed later in the Results section reflects the algorithms themselves rather than differences in data handling. This parallel structure also allows the cost and time efficiency analysis in the final stage to be evaluated independently of which classifier ultimately performs better, since the efficiency comparison is derived directly from the recorded cost and time attributes rather than from model output.

The dataset used in this study consists of 300 anonymized patient records compiled from several types of Indonesian health facilities, comprising Type A, Type B, and Type C hospitals, community health centers, and private clinics. It contains 22 attributes grouped into three categories. The first category covers thirteen clinical predictor attributes commonly used in cardiovascular risk assessment, namely age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic result, maximum heart rate achieved, exercise induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, the number of major vessels colored by fluoroscopy, and thalassemia status, a set of attributes consistent with those used in earlier heart disease classification studies (Alariyibi et al., 2023). The second category is the diagnosis label used as the classification target, which is evenly distributed between patients diagnosed with heart disease and patients without heart disease, thereby avoiding the class imbalance problem that can otherwise distort classifier performance. The third category consists of cost and time

related attributes, including the cost of conventional examination, the cost of data mining based early detection, the time required to obtain results through the conventional pathway, the time required through the data mining based pathway, the resulting cost difference, and the percentage of cost efficiency, which together form the basis of the economic analysis reported in the Results section. A complete description of all 22 dataset attributes, including their type and value range, is presented in Table 1.

Table 1 Description of Dataset Attributes (n = 300)

Category	Attribute	Type	Description / Range
Clinical predictor	Age	Numeric	29–77 years
Clinical predictor	Sex	Categorical	Male, Female
Clinical predictor	Chest pain type	Categorical	Typical angina, atypical angina, non anginal pain, asymptomatic
Clinical predictor	Resting blood pressure	Numeric	mmHg
Clinical predictor	Serum cholesterol	Numeric	mg/dl
Clinical predictor	Fasting blood sugar > 120 mg/dl	Categorical	Yes, no
Clinical predictor	Resting ECG result	Categorical	Normal, ST-T abnormality, left ventricular hypertrophy
Clinical predictor	Maximum heart rate achieved	Numeric	beats per minute
Clinical predictor	Exercise induced angina	Categorical	Yes, no
Clinical predictor	ST depression (oldpeak)	Numeric	0–6.2
Clinical predictor	Slope of peak exercise ST segment	Categorical	Upsloping, flat, downsloping
Clinical predictor	Number of major vessels	Numeric	0–3

Clinical predictor	Thalassemia	Categorical	Normal, fixed defect, reversible defect
Target label	Diagnosis	Categorical (binary)	Sakit Jantung (n = 150), Tidak Sakit Jantung (n = 150)
Cost and time	Conventional examination cost	Numeric	Rp977,000–Rp2,985,000
Cost and time	Data mining detection cost	Numeric	Rp120,000–Rp464,000
Cost and time	Conventional result time	Numeric	3–14 days
Cost and time	Data mining process time	Numeric	2–29 minutes
Cost and time	Cost difference	Numeric	Rp686,000–Rp2,739,000
Cost and time	Cost efficiency	Numeric (%)	68.91–93.69%
Contextual	Health facility type	Categorical	RS Tipe A, RS Tipe B, RS Tipe C, Puskesmas, Klinik Swasta

As summarized in Table 1, the thirteen clinical predictors combine both numeric measurements, such as resting blood pressure and maximum heart rate, and categorical indicators, such as chest pain type and thalassemia status, a mixed data structure that motivated the two distinct preprocessing paths described in the following paragraph.

An initial data quality check confirmed that the dataset contains no missing values across all 300 records, so no imputation procedure was required. Two parallel preprocessing paths were then applied to accommodate the different input requirements of the two classifiers used in this study. For the Naive Bayes model, all categorical clinical attributes, such as chest pain type and thalassemia status, were transformed into numeric form using label encoding, since the Gaussian variant of Naive Bayes applied in this study requires continuous numeric input. For the Decision Tree C4.5 model, categorical attributes were retained in their original textual form, consistent with the native handling of categorical splits in Quinlan's original C4.5 algorithm, while numeric attributes such as age and cholesterol level were kept as continuous values for threshold based splitting. In both paths, the dataset was partitioned into a training set of 240 records and a testing set of 60 records using an eighty to twenty ratio with stratified sampling to preserve the balanced proportion of diagnosis labels, and a fixed random seed of 42 was used throughout to make the partitioning identical and reproducible across both models.

The first classifier applied in this study is Gaussian Naive Bayes, a probabilistic classifier grounded in Bayes theorem that estimates the posterior probability of a class given a set of predictor attributes by assuming conditional independence among those attributes and a normal distribution for each continuous attribute. Despite this simplifying independence assumption rarely holding in real clinical data, Naive Bayes has

repeatedly been shown to perform competitively against more complex classifiers while requiring substantially less computational effort and training data, which makes it an attractive option for resource constrained health facilities (Rish, 2001). In this study, the model was implemented using the GaussianNB class from the scikit-learn library in Python, trained on the thirteen label encoded clinical attributes, and used to assign each patient the class with the highest posterior probability.

The second classifier applied is Decision Tree C4.5, an extension of the earlier ID3 algorithm developed by Quinlan that builds a tree structured classification model by recursively splitting the dataset on the attribute that yields the highest information gain ratio at each node, a refinement over simple information gain that reduces the algorithm's bias toward attributes with many distinct values (Quinlan, 1993). C4.5 further applies post pruning to the resulting tree in order to remove branches that contribute little predictive value, which helps prevent overfitting and produces a more interpretable set of diagnostic rules, an advantage that is particularly relevant for clinical decision support where transparency of reasoning is valued alongside raw predictive accuracy. In this study, the algorithm was implemented using the chefboost library in Python, which reproduces Quinlan's original gain ratio based procedure rather than the CART based decision tree implementation provided by default in scikit-learn. The model was trained on the same thirteen clinical attributes in their native categorical or numeric form and evaluated on the same training and testing partition used for the Naive Bayes model, ensuring that any difference in performance between the two algorithms can be attributed to the algorithms themselves rather than to differences in the underlying data split.

The predictive performance of both classifiers is evaluated using a confusion matrix that cross tabulates predicted and actual diagnosis labels on the testing set, from which four standard classification metrics are derived, namely accuracy, precision, recall, and F1 score. These four metrics were selected because they collectively capture both the overall correctness of a classifier and its balance between false positive and false negative errors, a balance that is especially important in a medical diagnosis context where the cost of missing an actual heart disease case differs from the cost of a false alarm (Sokolova & Lapalme, 2009). Accuracy is calculated as the proportion of correctly classified patients out of all patients evaluated, precision as the proportion of predicted heart disease cases that are truly positive, recall as the proportion of actual heart disease cases that are correctly identified, and F1 score as the harmonic mean of precision and recall, with the Sakit Jantung class consistently treated as the positive class across both models to ensure a fair comparison.

Beyond classification performance, this study analyzes the economic dimension of the proposed approach using the cost and time attributes recorded in the dataset. For each patient record, the difference between the conventional examination cost and the data mining based detection cost is calculated, and this difference is expressed as a percentage of the conventional cost to obtain a cost efficiency ratio, while the same comparison is applied to the time required to obtain a diagnostic result under each pathway. Descriptive statistics, including the mean, standard deviation, minimum, and maximum values of these efficiency ratios, are then computed both for the dataset as a whole and disaggregated by facility type, in order to examine whether the magnitude of cost and time efficiency varies across hospital tiers, community health centers, and private clinics, an approach consistent with prior findings that the real world value of an early detection strategy tends to depend on the specific healthcare context in which it is implemented (Oude Wolcherink et al., 2023).

RESULTS AND DISCUSSIONS

Result

This section presents the empirical findings addressing the two objectives outlined earlier, namely the classification performance of Naive Bayes and Decision Tree C4.5 in detecting heart disease, and the cost and time efficiency generated by the data mining based detection approach relative to conventional examination.

Classification Performance of Naive Bayes

The Naive Bayes model was trained on 240 patient records and tested on the remaining 60 records, evenly split between the two diagnosis classes. Table 2 presents the resulting confusion matrix on the testing set.

Table 2 Confusion Matrix of the Naive Bayes Model

Actual \ Predicted	Sakit Jantung	Tidak Sakit Jantung
Sakit Jantung	18	12
Tidak Sakit Jantung	10	20

Based on this confusion matrix, the Naive Bayes model achieved an accuracy of 63.33 percent, a precision of 64.29 percent, a recall of 60.00 percent, and an F1 score of 62.07 percent in identifying the Sakit Jantung class as the positive outcome of interest.

Classification Performance of Decision Tree C4.5

The Decision Tree C4.5 model was trained on the same 240 record training set and evaluated on the identical 60 record testing set used for the Naive Bayes model. Table 3 presents the resulting confusion matrix.

Table 3 Confusion Matrix of the Decision Tree C4.5 Model

Actual \ Predicted	Sakit Jantung	Tidak Sakit Jantung
Sakit Jantung	14	16
Tidak Sakit Jantung	11	19

The C4.5 model achieved an accuracy of 55.00 percent, a precision of 56.00 percent, a recall of 46.67 percent, and an F1 score of 50.91 percent.

3.3 Comparison of Classification Performance

Table 4 summarizes the four evaluation metrics for both classifiers on the same testing set.

Table 4 Comparison of Naive Bayes and Decision Tree C4.5 Performance

Metric	Naive Bayes	Decision Tree C4.5
Accuracy	63.33%	55.00%
Precision	64.29%	56.00%
Recall	60.00%	46.67%
F1 Score	62.07%	50.91%

Across all four metrics, the Naive Bayes model outperformed the Decision Tree C4.5 model on this dataset, with an advantage of 8.33 percentage points in accuracy and 11.16 percentage points in F1 score. To determine whether this observed advantage reflects a genuine difference in algorithm performance rather than variation attributable to a single train test split, a McNemar test was conducted on the paired predictions of both classifiers on the same 60 record testing set, a procedure recommended for comparing two classifiers evaluated on identical data (Dietterich, 1998). Table 5 presents the resulting contingency table.

Table 5 McNemar Contingency Table for Naive Bayes versus Decision Tree C4.5

	C4.5 Correct	C4.5 Incorrect
Naive Bayes Correct	23	15
Naive Bayes Incorrect	10	12

The continuity corrected McNemar chi square statistic was 0.64 with a p value of 0.424, which exceeds the conventional 0.05 significance threshold. This indicates that although Naive Bayes achieved higher point estimates across all four metrics, the difference in performance between the two classifiers on this testing set is not statistically significant, and should therefore be interpreted as a directional tendency rather than a conclusively superior algorithm.

Cost and Time Efficiency of Data Mining Based Early Detection

Table 6 presents the descriptive statistics of the cost and time attributes across all 300 patient records, comparing the conventional examination pathway with the data mining based early detection pathway.

Table 6 Descriptive Statistics of Cost and Time Efficiency (n = 300)

Variable	Mean	SD	Minimum	Maximum
Conventional examination cost (Rp)	1,853,393	359,846	977,000	2,985,000
Data mining detection cost (Rp)	270,360	59,316	120,000	464,000
Cost difference (Rp)	1,583,033	363,844	686,000	2,739,000
Cost efficiency (%)	84.82	4.66	68.91	93.69
Conventional result time (days)	8.26	3.39	3	14
Data mining process time (minutes)	15.18	8.19	2	29

At an aggregate level, the total conventional examination cost across all 300 patients amounted to Rp556,018,000, compared with Rp81,108,000 for the data mining based pathway, producing a total saving of Rp474,910,000, or an aggregate cost efficiency of 85.41 percent. This aggregate figure is calculated from the ratio of total saving to total conventional cost across the entire sample, and is therefore not identical to the 84.82 percent mean reported in Table 6, which instead represents the simple average of each patient's individual cost efficiency percentage. The two figures diverge slightly because patients with a larger conventional cost contribute more heavily to the aggregate ratio than to the simple average, though both consistently point to a cost saving in the range of 85 percent. In terms of time, the average conventional pathway required 8.26 days to produce a diagnostic result, whereas the data mining based pathway required an average of only 15.18 minutes.

Cost and Time Efficiency by Facility Type

Table 7 disaggregates the same cost and time variables by facility type.

Table 7 Mean Cost and Time Efficiency by Facility Type

Facility Type	n	Conventional Cost (Rp)	DM Cost (Rp)	Conventional Time (days)	DM Time (minutes)	Cost Efficiency (%)
RS Tipe A	42	1,781,119	283,524	9.12	14.57	83.28
RS Tipe B	69	1,870,855	266,942	8.48	15.54	85.19
RS Tipe C	55	1,863,382	267,127	7.91	15.42	85.26

Puskesmas	97	1,848,588	273,753	7.98	15.86	84.57
Klinik Swasta	37	1,900,622	257,703	8.11	13.05	85.86

Cost efficiency remains consistently high across all facility types, ranging narrowly between 83.28 percent at Type A hospitals and 85.86 percent at private clinics, a spread of only 2.58 percentage points.

Taken together, these findings indicate that while the classification performance of Naive Bayes and Decision Tree C4.5 differs numerically but not statistically on this dataset, the cost and time efficiency advantage of the data mining based early detection approach is substantial, statistically consistent in direction, and distributed evenly across every tier of the Indonesian healthcare system examined in this study, from tertiary hospitals down to community health centers and private clinics.

Discussion

The classification results reported in the previous section show that Naive Bayes achieved higher accuracy, precision, recall, and F1 score than Decision Tree C4.5 on this dataset, yet the McNemar test indicated that this numerical advantage was not statistically significant. This finding is instructive rather than disappointing, since it suggests that the choice between these two classifiers may matter less for diagnostic outcomes than the broader shift toward a data mining based diagnostic pathway itself. A plausible explanation for the relatively modest performance of C4.5 in this study lies in the categorical attributes with several possible values, such as chest pain type and thalassemia status, which under the gain ratio splitting criterion can produce branches supported by comparatively few training instances once the tree grows past its first two or three levels, a tendency that becomes more pronounced on a moderately sized dataset of 300 records. Gaussian Naive Bayes, by contrast, estimates a single mean and variance per attribute per class regardless of how many categories that attribute contains, which may explain its comparatively more stable performance on a dataset of this size. This pattern echoes earlier comparative work showing that classification accuracy for heart disease prediction can vary considerably across algorithms and is sensitive to the specific dataset used, rather than following a fixed ranking that holds universally (Alariyibi et al., 2023).

The moderate accuracy achieved by both classifiers, 63.33 percent for Naive Bayes and 55.00 percent for C4.5, also warrants direct acknowledgment rather than being understated. These figures are lower than the accuracy levels reported in several previous heart disease classification studies using ensemble or hybrid approaches on larger benchmark datasets (Latha & Jeeva, 2019), which suggests two non exclusive explanations. First, the sample size of 300 records, while sufficient to illustrate the intended cost and time efficiency analysis, is relatively small for a thirteen attribute classification task, and second, single algorithm classifiers such as standalone Naive Bayes and standalone C4.5 without ensemble refinement tend to leave some predictive accuracy on the table compared with boosted or bagged variants. Rather than overstating the diagnostic reliability of either model, this study positions Naive Bayes and C4.5 as a baseline comparison intended primarily to support the cost and time efficiency analysis, and explicitly leaves the pursuit of higher diagnostic accuracy through ensemble methods as a direction for future work.

In contrast to the modest and statistically inconclusive difference in classification performance, the cost and time efficiency findings in this study are both substantial and remarkably consistent. An average cost efficiency of 84.82 percent per patient, an aggregate efficiency of 85.41 percent across the full sample, and a reduction in diagnostic

turnaround from an average of 8.26 days to roughly 15 minutes together represent a large and economically meaningful shift regardless of which classifier ultimately governs the diagnostic decision. This result is consistent with the broader health economic literature, which has found that early cardiovascular detection strategies tend to be cost effective relative to delayed or conventional detection, while also noting that the magnitude of that value depends heavily on the specific healthcare context in which it is measured (Oude Wolcherink et al., 2023). The present study extends that literature by providing a context specific quantification for Indonesia, a gap previously noted in the artificial intelligence and cardiovascular health literature as lacking empirical grounding outside high income healthcare systems (Ahmed et al., 2025). Given that heart disease already represents the single largest catastrophic cost burden financed through Indonesia's national health insurance program, an efficiency gain in this range, if realized at scale, would carry direct relevance for national health financing sustainability rather than remaining a purely academic result.

The consistency of cost efficiency across facility types, ranging narrowly from 83.28 percent at Type A hospitals to 85.86 percent at private clinics, further suggests that the economic benefit of a data mining based early detection approach is not contingent on the resource level of the facility implementing it. This is a particularly encouraging signal for the Indonesian context, where Puskesmas accounted for the largest share of patients in this dataset yet operates with comparatively limited diagnostic infrastructure relative to tertiary hospitals. If this pattern holds in real world deployment, it implies that digital transformation initiatives built around lightweight classification models need not be reserved for well resourced hospitals, but could plausibly be extended to primary care level facilities where the burden of diagnostic delay and cost is often felt most acutely by patients.

Several limitations of this study should nonetheless be acknowledged. The dataset used, while structured consistently with established heart disease classification attributes, consists of 300 records collected and encoded for this study, and the moderate classification accuracy obtained suggests that the relationship between clinical attributes and diagnosis in this dataset is noisier than in some widely used benchmark datasets, a factor that should be considered when interpreting the generalizability of the classification results. The evaluation was also based on a single stratified train test split rather than repeated k fold cross validation, which, although adequate for the exploratory comparison conducted here, means that the reported performance figures carry some sampling variability that a cross validated design would help quantify more precisely. In addition, the cost figures used in the efficiency analysis, while structured to reflect realistic ranges across facility types, should ideally be validated against official tariff schedules from Indonesia's Ministry of Health or BPJS Kesehatan before being used to inform actual policy or budgeting decisions. Finally, this study evaluated only two classical classifiers, and did not benchmark them against more recent ensemble or deep learning approaches that have shown improved accuracy in related work.

These limitations point toward several directions for future research. Subsequent studies could apply k fold cross validation and repeated resampling to obtain more stable performance estimates, and extend the comparison to ensemble methods such as Random Forest or gradient boosted trees to examine whether higher classification accuracy can be achieved without sacrificing the interpretability that motivated the use of C4.5 in this study. A comparable strategy applied to the Cleveland and IEEE Dataport heart disease datasets found that combining hyperparameter optimization with five fold cross validation and a soft voting ensemble across six classifiers, including Naive Bayes,

raised accuracy above 93 percent, illustrating the scale of improvement such refinements could plausibly offer if applied to the present dataset (Chandrasekhar & Peddakrishna, 2023). Future work should also validate the cost model against official Indonesian healthcare tariff data collected prospectively across a larger and more diverse set of facilities.

From an information systems perspective, a natural next step would be to translate the classification and efficiency framework developed here into a functional clinical decision support prototype, integrating the trained models into a simple web based or electronic health record embedded interface that frontline health workers at Puskesmas or private clinics could use directly. This direction aligns with evidence from other developing country contexts, where accurate decision support systems have been shown to play a vital role in the early detection of heart disease specifically in remote, semi urban, and rural areas that lack access to specialist cardiologists (Rani et al., 2021). Moving this line of research from an offline comparative analysis toward a deployable digital health tool would therefore be consistent with the broader digital transformation agenda referenced at the outset of this article.

CONCLUSION

This study set out to build and compare Naive Bayes and Decision Tree C4.5 models for the early detection of heart disease and to quantify the cost and time efficiency generated by a data mining based detection approach relative to conventional examination, using clinical and financial data drawn from multiple types of Indonesian health facilities. The Naive Bayes model achieved higher classification performance than Decision Tree C4.5, with an accuracy of 63.33 percent compared with 55.00 percent, although a McNemar test confirmed that this difference was not statistically significant, indicating that neither algorithm can be considered conclusively superior to the other on this dataset. The more decisive finding of this study concerns economic efficiency, where the data mining based detection pathway produced an average cost saving of 84.82 percent per patient and reduced diagnostic turnaround from an average of 8.26 days to approximately 15 minutes, with this efficiency remaining consistently high across every facility type examined, from tertiary hospitals to community health centers and private clinics.

These results support the conclusion that the primary value of a data mining based approach to heart disease early detection lies less in outperforming a single alternative classification algorithm and more in the substantial and broadly distributed reduction in cost and time it offers relative to conventional diagnostic procedures. Given that heart disease already represents the largest catastrophic cost burden financed through Indonesia's national health insurance program, this efficiency gain carries practical relevance beyond the immediate scope of this study, particularly for primary care facilities such as Puskesmas that serve the largest patient volumes with comparatively limited resources.

This study contributes to the health informatics literature by providing an empirically grounded, Indonesia specific quantification of the economic value of data mining based cardiovascular screening, an area of evidence previously identified as limited outside high income healthcare systems. At the same time, the moderate classification accuracy obtained, the use of a single train test split, and the reliance on a dataset of 300 records constitute clear limitations that should temper any claim of diagnostic readiness for clinical deployment. Future research is therefore encouraged to validate these findings using cross validated evaluation on larger and more diverse

datasets, to benchmark ensemble based classifiers against the baseline models used here, and to extend this analytical framework into a functional clinical decision support prototype that can be tested directly within Indonesian primary care settings as part of the broader digital transformation of the national health system.

REFERENCES

Ahmed, A., Khaled, A., Waqar, M., Hashmi, D. J. A., Alfanash, H. A., Almagharbeh, W. T., Hamdache, A., & Elmouki, I. (2025). AI-Driven Early Detection of Cardiovascular Diseases: Reducing Healthcare Costs and improving patient Outcomes. *South Eastern European Journal of Public Health*, 3591–3601. https://doi.org/10.70135/seejph.vi.3521

Aisyah, D. N., Setiawan, A. H., Lokopessy, A. F., Faradiba, N., Setiaji, S., Manikam, L., & Kozlakidis, Z. (2024). The Information and Communication Technology Maturity Assessment at Primary Health Care Services Across 9 Provinces in Indonesia: Evaluation Study. *JMIR Medical Informatics*, *12*(1), e55959. https://doi.org/10.2196/55959

Alariyibi, A., El-Jarai, M., & Maatuk, A. (2023). Evaluating The Accuracy of Classification Algorithms for Detecting Heart Disease Risk. *Machine Learning and Applications: An International Journal*, *10*(4), 01–12. https://doi.org/10.5121/mlaij.2023.10401

Ananey-Obiri, D., & Sarku, E. (2020). Predicting the Presence of Heart Diseases using Comparative Data Mining and Machine Learning Algorithms. *International Journal of Computer Applications*, *176*(11), 17–21. https://ijcaonline.org/archives/volume176/number11/31245-2020920034/

Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. *Processes*, *11*(4), 1210. https://doi.org/10.3390/pr11041210

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Harian Kompas. (2024). *Beban Ekonomi Penyakit Jantung Rp 67,34 Triliun*. PT Kompas Media Nusantara. https://kompas.id

Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, *16*, 100203. https://doi.org/10.1016/j.imu.2019.100203

Lindstrom, M., DeCleene, N., Dorsey, H., Fuster, V., Johnson, C. O., LeGrand, K. E., Mensah, G. A., Razo, C., Stark, B., Varieur Turco, J., & Roth, G. A. (2022). Global Burden of Cardiovascular Diseases and Risks Collaboration, 1990-2021. *JACC*, *80*(25), 2372–2425. https://doi.org/10.1016/j.jacc.2022.11.001

Mulure, N., Hewadmal, H., Khan, Z., Mulure, N., Hewadmal, H., & Khan, Z. (2024). Assessing Barriers to Primary Prevention of Cardiovascular Diseases in Low and Middle-Income Countries: A Systematic Review. *Cureus*, *16*. https://doi.org/10.7759/cureus.65516

Oude Wolcherink, M. J., Behr, C. M., Pouwels, X. G. L. V., Doggen, C. J. M., & Koffijberg, H. (2023). Health Economic Research Assessing the Value of Early Detection of

Cardiovascular Disease: A Systematic Review. *PharmacoEconomics*, *41*(10), 1183–1203. https://doi.org/10.1007/s40273-023-01287-2

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers. https://google.com

Rish, I. (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, *3*, 41–46. https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier

Shirzad, E., Ataei, G., & Saadatfar, H. (2021). Applications of data mining in healthcare area: A survey. *Engineering and Applied Science Research*, *48*(3), 314–323. https://doi.org/10.14456/easr.2021.34

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–40. https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining

World Health Organization. (2025). Cardiovascular diseases (CVDs) Fact Sheets. *WHO News Room Fact Sheets*. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Yadav, S., Singh, D. M. K., & Kumar, P. (2023). Data Mining Applications for Enhancing Healthcare Services: A Comprehensive Review. *International Journal of Engineering Technology and Management Sciences*, *7*(5), 325–333. https://doi.org/10.46647/ijetms.2023.v07i05.038