EasyChair Preprint

№ 14927

Comparative Analysis of Community Rating and Dynamic Pricing on Enhancing Healthcare Accessibility in Low-Income Countries: a Case Study of Tanzania

Godfrey Justo and Fadhili Meena

September 18, 2024

Comparative Analysis of Community Rating and Dynamic Pricing on Enhancing Healthcare Accessibility in Low-Income Countries: A Case Study of Tanzania

Godfrey N Justo[1] and Fadhili Z Meena[2]
[12] University of Dar es Salaam, College of Information and Communication Technologies,
P.O BOX 33335, Dar es Salaam, Tanzania.
njulumi@gmail.com, fadhilimeena@gmail.com

Abstract. Over the last two decades out-of-pocket (OOP) payments and external donations have contributed over 60% of current health expenditure (CHE) in low-income countries (LICs), in which OOP account for more than 40%, indicating a heavy reliance on individual contributions. Current studies on pricing models for healthcare largely focuses on achieving computational efficiency, but not the pricing effect on healthcare accessibility that influence healthcare improvement. Analysis of healthcare pricing models based on cost affordability can bridge this gap. Commonly used pricing models such as community rating (CR), dynamic pricing (DP) and OOP are considered. DP is a machine learning (ML) model based on Tanzania's National Panel Survey (NPS) data, while the CR model is based on rates from the Tanzania's Act Supplement for the mandatory public health insurance scheme of 2023, to the Ministry of Health (MoH). A pure premium approach for the DP model and current rates for the CR model is employed for comparative purpose. The results showed that CR does not significantly improve healthcare costs compared to DP (p-value = 0), conversely, DP significantly outperforms CR with p-value 3.49×10^{-08}. Moreover, DP model remains superior to CR until loading factor range of 5.7 and 6.4, where no significant difference, beyond which DP increases healthcare costs. Likewise, DP outperforms OOP until the loading factor range of 0.1 and 0.2, where costs are insignificantly different, above which DP increases costs. Load factor analysis confirm DP to significantly enhance healthcare accessibility by reducing cost compared to CR and OOP pricing models.

Keywords: Healthcare accessibility, Community rating, Dynamic pricing, Pricing affordability, Individual contribution, Machine learning

1 Introduction

1.1 Background

Current healthcare financing research focuses on computational reliability and the application of advanced AI models [1,2,3], for accuracy, albeit pay no attention pricing

effect on enhancement of healthcare access, a fundamental goal for sustainable human development (SDG) and Universal Healthcare Coverage (UHC) [4]. Effective execution of healthcare access policies, such as the UHC requires efficient funding, especially for disadvantaged groups [5], which is a notable challenge in low-income countries (LICs) such as Tanzania, where over 45% of current health expenditures (CHE) rely on out-of-pocket payments (OOP), compared to high-income countries where more than 70% come from government budgets [6]. The recent Global health expenditure database (GHED), shows the trend has persisted for over 20 consecutive years to 2022 [7]. As such due to limited sources of internal revenue, individual contributions through OOP remain a pillar for financing CHE in LICs [8].

Community rating (CR) is a prevalent approach for estimating individual contributions in LICs to major healthcare protection schemes. The CR is characterized by similar contribution rates [9], albeit, indexed or categorized based on various factors such as income and geographical regions [10]. Conversely, dynamic pricing (DP) is based on mathematical models that account for various individual attributes (including risk factors such as age, gender, health history, and healthcare utilization pattern) and evaluate how they change over time to make contribution estimates that reflect expected individual healthcare costs. Recent research has leveraged Machine Learning (ML) models to enhance DP approaches due to their computational efficiency and ability on diverse datasets [11,12,13].

Pricing affordability is one of important aspects in Levesque's framework for healthcare access measures [14]. However, the relative effect of pricing models on enhancing healthcare access remains unexplained by current research. This study seeks to bridge the gap by analyzing the effect of OOP, CR and DP pricing models in improving healthcare accessibility, considering the case of Tanzania.

Ensuring access to quality healthcare without financial hardship is crucial for sustainable development, particularly in low-income countries where individuals often fund their own care. Pricing methods like CR and DP have been proposed for estimating premium contributions. However, the current research attention has been on computational efficiency, but it remains unclear on the effect of such pricing models to improve healthcare accessibility. Understanding the relative effect is important for effective implementation of healthcare access policies, such as the UHC. This study is guided by the following research questions:

- What are the key factors affecting healthcare pricing for accessibility with respect to CR and DP models?
- How significant is the difference between CR and DP models in healthcare access improvement through affordability?

The study focuses on low-income countries, with Tanzania as the case study. The study scope is limited to a comparative analysis of commonly used financing approaches: OOP, CR and DP models. Their relative pricing effect in influencing individual preferences and healthcare access is subsequently assessed based on price estimates from each model. The outcomes are expected to inform healthcare access policies and programs, such as UHC, in similar contexts. Clarifying the strengths and weaknesses of a financing model, helps to guides decision-makers toward more effective

approaches that balance financial sustainability along with the broader goal and more equitable healthcare access.

1.2 Materials and Methods

The research design employs a quantitative approach, using descriptive and predictive analysis of numerical data related to healthcare utilization [15]. The case study design enables an in-depth examination of healthcare financing systems. Data is collected from secondary sources, including National Pannel Survey (NPS), Demographic Health Survey (DHS) [16,17], and GHED [7]. The GHED provides national and global health expenditure data, offering insights into health financing patterns and supporting generalizing the study to LMICs. The DHS is based on household surveys in low- and middle-income countries, including Tanzania, gathering data on health indicators, healthcare utilization, insurance uptake, fertility, maternal, and child health, providing valuable insights into Tanzania's healthcare landscape [17]. The NPS collects longitudinal data on healthcare utilization, expenditure patterns, and demographics from 2008-2021, capturing trends and dynamics in individual healthcare costs [16], as summarized in Table 1.

Table 1: Summary of data collected from NPS

Data Collection Wave	Number of Observations	Number of Attributes
Wave 2 2010 - 2011	20559	42
Wave 3 2012 - 2013	25412	42
Wave 4 2014 - 2015	16285	43
Wave 5 2020 - 2021	23592	43

The methods in [18] for sample design for unknown populations is used for sample design that target a small but representative sample size, which suggests an optimal sample size of 385 for a large population at a 95% confidence level. In the NPS repository context, the sample size for each wave exceeds 10,000, which is sufficient for the models. The resampling technique, Small Minority Oversampling (SMO), is used to balance biased samples. The key attributes include age, gender, healthcare service frequency, and claim size [12,19]. Python version 3.11.4 is used for exploratory data analysis, model training, model evaluation and hypothesis testing.

1.2.1 Exploratory Data Analysis

Different tests are conducted to explore the nature of attributes in the data set including the chi-square test for independence on categorical variables like gender and healthcare utilization status, and Pearson’s correlation coefficient test across continuous (numerical) attributes such as age [20,21]. Histograms are used on univariate EDA, to visually identify the distribution of individual continuous variables and bar plots to identify the distribution of individual categorical variables. Further, test for the goodness of fit is used to identify the best-fit distribution for each of the variables, a crucial step for further simulation design to create more comprehensive data to train the ML

model for understanding the best distribution of attribute that is reproducible independent of the raw data [22].

Missing values within the dataset challenge the accuracy of statistical analyses and model performance. To address this issue, the extent of missingness are identified across variables and the cause of missing. In turn, a multistage predictive model replaced missing values with estimated values derived from existing data patterns [23]. For systematic missing, for example, if a person is aged 5 years and has an NA marital status value, we directly feed 'Never Married'. A multivariate random forest regressor (classifier for categorical variables) fills the remaining missing observations, provided the proportion of missingness is less than 0.7, otherwise, the attribute is dropped.

The standard deviation (Z-score) analysis method is utilized to identify outliers across variables. Once identified, outliers are winsorized, truncated, or excluded from analysis, depending on the nature of the data [24,25]. For a variable X;

$$z - score\ (Z) = \frac{x - E[X]}{\sqrt{var(X)}} \quad \text{and } P(Z < 3) \approx 99.87\%$$

hence, observations above 3 standard deviations are considered as outliers; for the target variable, outliers are dropped, while for other variables, a predictive model is used to test how the observation with outliers can be explained by other covariates. That is the predicted values and recorded outliers are compared based on Mean Absolute Percentage Error (MAPE). Observations from variables with MAPE above 50% of the predicted values are considered; otherwise, the recorded observations are retained.

The handling of imbalanced categorical attributes employed the Synthetic Minority Over-sampling Technique (SMOTE) to rebalance class distributions [26,27,28]. When observed that the resampling of one variable caused an imbalance of other categorical variables, the tree-based SMOTE is adopted rather than just focusing on a single categorical variable, to ensure that the target variable is balanced while other categorical are less biased. As data scales varied rapidly, especially on variables with many categories, only variables with two categories are rebalanced. Further, a statistical test on improved variability of class proportions among categorical variables is performed. That is, if $X = \{x_1, x_2 \dots k\}$ is the variance of proportions of classes in set of all categorical variables $1,2,..$ before class balance and $Y = \{y_1, y_2 \dots k\}$ is the variance of proportions of classes in set of all categorical variables $1,2,..$ after class balance; it is hypothesized that *class balance approach reduce the overall variance among the categorical variables*, to which; $\hat{\sigma}_Y^2 < \hat{\sigma}_X^2$

$$\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \sim F_{0.05,k-1,k-1}$$

where k is the number of categorical variables involved in class balance. Though, should be all categorical variables, but for computation constraint, only four variables are included. In addition, an F-test is performed, with a significance level of 5% and on a set of p-value. With confident the adopted class balance approach significantly reduced the overall class proportion variance among all categorical variables (with binary classes) in the data.

1.2.2 Feature Selection and model development

The categorical variables are transformed into numerical formats while preserving their semantic meaning by Label encoding technique. Unlike other approaches, such as one hot encoding, the Label encoding technique can address the dimensionality issue for high cardinality categorical variables [29,30]. Leveraging on outliers handling the feature scaling techniques are performed by a standardization approach. To identify relationships between features and the target variable (individual healthcare contribution), correlation analysis is conducted, in which the features with high correlation coefficients are retained for model training, while those with low correlation are excluded to reduce dimensionality and multicollinearity. The Bartlett's test of sphericity (BTS) and Kaiser-Meyer-Olkin (KMO) test of sampling adequacy is employed to evaluate if the correlation between the factors is significantly different from zero, more so if the data is suitable for factor analysis [31,32]. The BTS evaluated if observed correlation matrix is significantly different from identity matrix (that is variables are not intercorrelated). The factor analysis is employed if the correlation matrix is significantly different from identity matrix. Under this scenario test statistic T is

$$T = -\left(N - 1 - \frac{2p+5}{6}\right) \times \ln(|R|)$$

where, N is sample size, p is number of variables, $|R|$ is the determinant of the correlation matrix $T \sim X^2_{\left(\frac{p(p-1)}{2}\right)}$. On the other hand, KMO is used to measure the adequacy of the sample for factor analysis, and the KMO of 0.6 overall is acceptable. The KMO statistic can be given as $KMO_j = \frac{\sum_{i \neq j} R_{ij}^2}{\sum_{i \neq j} R_{ij}^2 + \sum_{i \neq j} U_{ij}^2}$ whereby R_{ij}^2 is the correlation matrix and U_{ij}^2 is the partial covariance matrix [31].

A tree-based algorithm (Random Forest) is used to assess feature importance using the Gini criterion. Features with higher Gini scores are prioritized for inclusion in the model, while less important features may be excluded to simplify the model and improve interpretability. Furthermore, the permutation test is used to identify a set of features that significantly contribute to the model's predictive performance, to improve parsimony and reduce overfitting. Principal Component Analysis (PCA) is used to identify latent variables that capture the most important information in the data to reduce computational complexity.

The model development involved three stages; the prediction of healthcare cost (continuous variable) by regression-based model, the prediction of the chance for filling the claim based on the probabilistic classification model, where a pure premium principle for estimating the individual's premiums is employed. The regression model is developed using a set of tree-based regression algorithms, including Random Forest (RF), Gradient Boosting (GB) and Extreme gradient boosting (XGB) algorithms due to their proven effectiveness in handling complex, nonlinear relationships within the data [33]. Model training made use of historical healthcare expenditure data from Tanzania's

NPS, constitute of longitudinal dataset with detailed information on individual healthcare utilization, demographics, and expenditure patterns, making it an ideal source for training a predictive model. The dataset is split into training and validation sets to ensure the robustness and generalizability of models.

The randomized search cross-validation is employed during hyper-parameter optimization to find the best hyper-parameters for the RF and GB tree-based regression models. This method efficiently explores the hyper-parameter space by sampling from specified distributions, balancing computational resources, and improving model accuracy. The model selection step assessed the performance of regression and classification algorithms tailored to predicting healthcare costs and the probability of filing a claim. For the regression task, the mean squared error (MSE), root mean square error (RMSE), and mean absolute error (MAE) served as primary indicators of predictive accuracy (i.e., lower value of a metric signifies closer alignment between the model's predictions and the actual healthcare costs, reflecting superior performance). The goodness of fit is evaluated by the coefficient of determination (R-squared), which quantifies the proportion of variance in healthcare costs explained by the model's predictors (i.e., higher R-squared values indicate a better fit of the regression model to the data, suggesting its efficacy in capturing underlying relationships). The accuracy, precision, recall, and F1-score are employed for the classification task of predicting claim probabilities. Accuracy gauges the correctness of the model's predictions, while precision measures the proportion of true positives among all positive predictions. Recall, or sensitivity, quantifies the model's ability to identify true positives correctly. The F1-score balances precision and recall, comprehensively evaluating classification performance. Moreover, the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC-ROC) are also considered to visualize the trade-off between sensitivity and specificity. Ultimately, the model selection process prioritizes the algorithm demonstrating superior performance across the metrics, ensuring robustness and generalizability in predicting healthcare costs and claim probabilities.

The individual contribution for ML based DP model is estimated as follows: If Y_i is the random output healthcare cost for individual i form the regression model, and p_i is the probability for a particular individual to make a claim. Based on pure premium, the principal premium equals the expected loss.

$$\pi_i = E[Y_i] = y_i \times p_i$$

where π_i is the pure premium for i^{th} individual. This principle is employed to ignore external expenses such as administrative costs. However, a sensitivity analysis is conducted by introducing the loading factor in the pure premium principle to examine how the proportionate change in external expenses affects the inference. When adjusted to loading factor θ which account for cost out of the sum assured (can include underwriting expenses, management costs, and claim settlement expenses), then

$$\pi_i = (1 + \theta)E[Y_i] = (1 + \theta) \times y_i \times p_i$$

Individual contribution for the community rating approach is based on the Act Supplement for the mandatory public health insurance scheme of 2023, its respective bill of August 2022, and the related information to the public from the Ministry of Health (MoH) [34,35,36,37], as summarized in Table 2.

Table 2: Contribution rates under the CR approach

Condition	Contribution Estimate ($\hat{P}$)
If individual is from formal sector	$\hat{P} = \frac{6}{100} \times S$ where S is the gross annual salary of an individual
If individual is from informal sector	$\hat{P} = 84{,}000$
Group (family) of 6	$6\hat{P} = 340{,}000$

The healthcare accessibility is evaluated based on the two pricing approaches: Let price estimates from the DP and CR models be $\hat{P}_i$ and $\hat{P}_{m_i}$, respectively. Recall from the theory of expected utility, an individual's attitude (behavior) towards risk is geared towards increasing their expected utility, and the opportunity cost principle, where individuals decide to reduce money lost by selecting one option over the other. A hypothesis is formulated to test for two samples to evaluate whether a significant price difference causes people to prefer one pricing approach over the other:

$$\boldsymbol{H_0 : \mu_1 = \mu_2}$$
$$\boldsymbol{H_1\ \mu_1 \neq \mu_2}$$

Given n the sample size of the latest NPS wave (test) data, where μ_1- average individual contributions for the population if exposed to community rating approach and μ_2- average individual contributions for the population if exposed to a dynamic pricing approach. Since one sample is evaluated over two different approaches, an independent two-sample test is used [38]. The test statistic T

$$\boldsymbol{\bar{x}_1 = \overline{(\hat{P}_\iota)} = \frac{1}{n}\sum_{i=1}^{n} \hat{P}_i}$$

$\bar{x}_2 = \overline{(\hat{P}_{m\iota})} = \frac{1}{n}\sum_{i=1}^{n} \hat{P}_{mi}$

$s_1^2 = var(\hat{P}_i) = \frac{1}{n-1}\sum_{i=1}^{n} \left(\hat{P}_i - \overline{(\hat{P}_\iota)}\right)^2$

$s_2^2 = var(\hat{P}_{mi}) = \frac{1}{n-1}\sum_{i=1}^{n} \left(\hat{P}_{mi} - \overline{(\hat{P}_{m\iota})}\right)^2$

$s_p = \sqrt{(s_1^2 + s_2^2) \times \left(\frac{n-1}{2n-1}\right)}$

$T = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{2}{n}}}$

with probability values $p - values = 2 * \Theta(\mathrm{T})$ at different values of $\alpha = 0.05, 0.01, 0.005$ is evaluated to identify if there is a significant difference in the individual contributions, which consequently favor one pricing approach over the other.

2 Results

2.1 Exploratory factor analysis

The exploratory factor analysis (FA), shows continuous variables are somewhat related to the target variable. However, it is insufficient to conclude that the correlations significantly differ from zero. Thus, it prompts to use of BTS, and for the correlated variables the KMO is employed to explore the number of latent variables. The KMO test result showed average of 0.74698 and 0.820831 in general, indicating with confident that the data is suitable for FA, since at least KMO of 0.6 is required in general. The scree plot in Figure 1 is subsequently used to determine number of variables to retain from the data based on the magnitude of the eigenvectors, where ten variables with the eigen-values greater than one are observed to meet selection criteria.

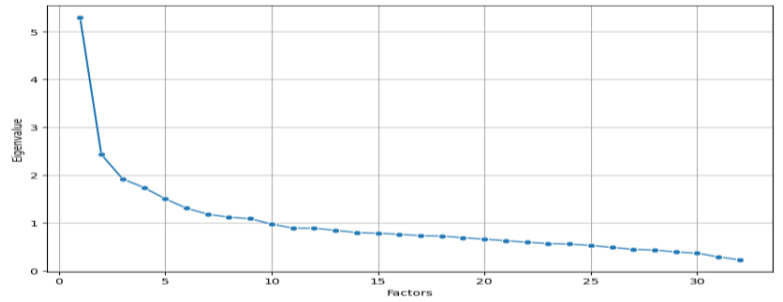

Figure 1: Eigenvectors magnitude on data

A linear regression model is fitted to asses how variables are best suited for predicting healthcare costs. The regression analysis results indicate a model with poor explanatory power, as evidenced by the low R-squared value of 0.07. The F-statistic is high at 2927.00, indicating a statistically significant relationship between the independent and dependent variables despite the low R-squared. The Omnibus test, Jarque-Bera test, and their associated p-values suggest that the residuals are not normally distributed, with high skewness (18.56) and kurtosis (692.43), indicating significant departures from normality. Moreover, the extremely high condition number suggests potential issues with multicollinearity that prompt to employ PCA. The Durbin-Watson statistic of 1.858 suggests the presence of some positive autocorrelation in the residuals. Overall, while the model appears to be statistically significant, its practical utility is limited due to the poor fit, non-normality of residuals, potential multicollinearity, and autocorrelation, generally suggesting all variables are significant predictors at 5% but have low power to linear model. This prompt to employ Random Forest regressor and Gradient Boosting algorithms.

A logit model is fitted to evaluate significant factors that influence the chance of an individual to incur a particular cost (i.e., predicting healthcare use). The logistic regression model reveals several important findings regarding the relationship between the predictor variables and the binary outcome variable. The model exhibits a pseudo-R-squared value of 0.3145, indicating moderate predictive power. The coefficients for each predictor variable provide insight into their respective effects on the log odds of

the outcome. For instance, some variables portray positive coefficients suggesting that an increase in their values is associated with an increased likelihood of the outcome. Conversely, other variables portray negative coefficients, indicating a decrease in the probability of the outcome as their values increase. Notably, the p-values associated with each coefficient are all very low, indicating that most predictor variables are statistically significant in predicting the outcome. However, a few variables had relatively high p-values, suggesting that they may not be significant predictors of the outcome. This signals potential for elimination if their information value is lower when evaluated with the feature importance Gini criterion of the RF model.

2.2 Models Development

Different models are developed including a regression model for the prediction of healthcare cost, a classification model for predicting the probability of a person incurring a particular cost and a pure premium principle used for estimation of individual contribution based on the dynamic model. Moreover, a community rating model, is developed using a rule-based approach from the ratings outlined in Table 2.

2.2.1 Regression Model

Recall that all variables' correlations are significantly different from zero and that all variables are significant predictors. However, some multicollinearity exists. The PCA is performed to address the latter. The dataset is separated into two sections: model development data (NPS wave 2 - 4) and model validation data (NPS wave 5). Model development data is divided into two sets (training and testing set). The model is first trained to determine the importance of the features using the Gini information criterion as depicted in Figure 2.

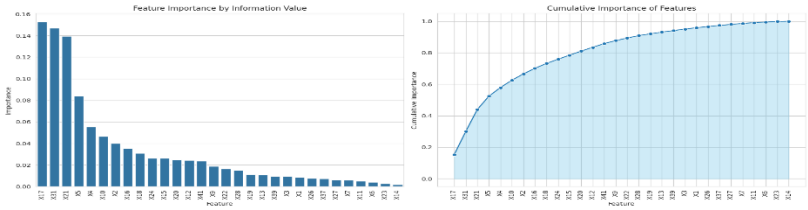

Figure 2: Feature importance for the regression model

The results in Figure 2, support the inference from Section 2.1 that all variables are important predictors for the regression model. In addition, more than 86% of the variables are required to explain at least 90% of the information for the target variable. Consequently, all variables are selected for the regression model and the observed multicollinearity addressed with PCA. The PCA in Figure 3, show that more than 76% of components are required to explain 90% of the variance. Hence, dimensionality reduction is employed by using PCA and retain only 23 components for model development.

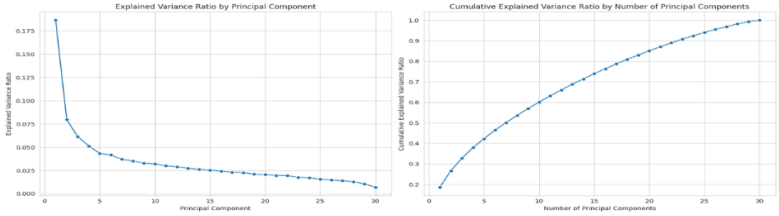

Figure 3**:** PCA variance analysis for the regression model

Models are trained on raw data (i.e., without rescaling or dimensionality reduction), and subsequently with rescaled data first by standard scaling approach on the training data only to prevent information leakage, and by PCA with only 23 components retained. Due to limited computational resource, iterative hyper-parameter optimization is not performed at start rather random initialization. Models are trained by the RF regressor, GB and XGB algorithms, as depicted in Table 3.

Table 3: Regression model performance before (shaded rows) and after scaling/PCA

Model	Train R-Squared	Test R-Squared	RMSE	Computation Time (sec)
Random forest (RF)	0.981437	0.822853	15520.307978	420
	0.97278	0.75403	18288.2958	1080
Gradient boosting (GB)	0.624619	0.5582952	24510.03213	60
	0.6031127	0.5402546	25003.0084	240
Extreme gradient boosting (XGB)	0.915697	0.7828987	17181.745050	3
	0.9012672	0.7174	19600.39311	15

Considering RMSE, the RF regressor outperforms XGB by about 7%. On the other hand, considering the computation time, XGB has more than 7100% lower computational time compared to the RF regressor. Hence, the XGB regressor without scaling and dimensionality reduction is adopted for the prediction of healthcare costs.

2.2.2 Classification Model

Recall from Section 2.1 that all variables' correlations are significantly different from zero, and that variables which are not significant predictors are dropped. However, some multicollinearity exists in which the PCA is used. As in Section 2.2.1, the dataset is separated on similar basis and first train the model to determine the importance of features using the Gini information criterion as depicted in Figure 4

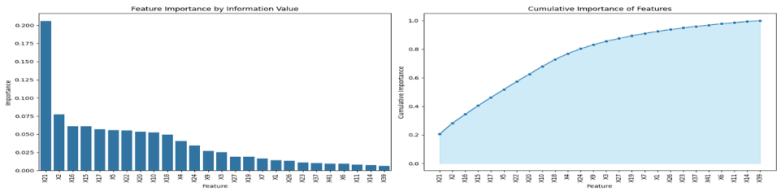

Figure 4: Feature importance for the classification model

The results, in Figure 4, support the inference in Section 2.1 that all variables are important predictors for the classification model. In addition, only 77% of the variables are required to explain at least 90% of the information for the target variable. This prompts to take all variables for the classification model and proceed to address the observed multicollinearity with PCA in Figure 5.

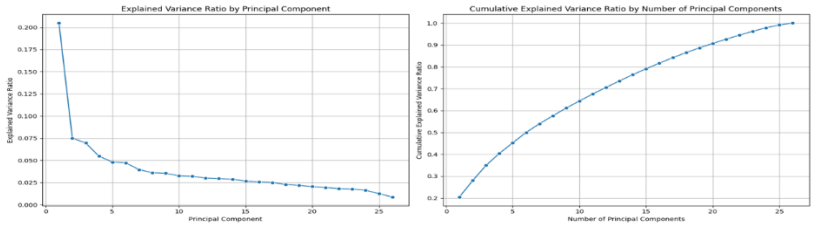

Figure 5: PCA variance analysis for the classification model

The PCA, in **Error! Reference source not found.**, shows that more than 76% of c omponents are required to explain 90% of the variance. Hence, dimensionality reduction is employed through PCA and only 20 components are retained for model development. Table 4, shows the classification model performance before and after scaling and PCA.

Table 4: Classification model performance before (shaded row) and after scaling/ PCA

Model	Train Accuracy	Test Accuracy	Computation Time (sec)
Random forest (RF)	0.99978	0.880618	207
	0.9999654	0.848723	381
Gradient boosting (GB)	0.7212755	0.7168279	245
	0.67244438	0.66416666	485
Extreme gradient boosting (XGB)	0.834519982	0.807782258	48.7
	0.7920771	0.7470027	4

The bias in RF is larger compared to the XGB classifier, so for the computational time, the XGB classifier without scaling and dimensionality reduction pipeline is adopted.

2.2.3 Contribution estimation

The dynamic model employs pure premium principle to find individual contribution, in which the cost is estimated by XGB regressor and probability of incurring the cost is estimated by the probabilistic XGB classifier. Individuals whose contributions are above 3 standard deviations of the whole sample individual contributions are classified as high-risk individuals, while others are normal-risk individuals. General individual contribution distributions follow beta distribution similar to individuals with normal risk, while high-risk individuals follow Pareto distribution with parameters as described in

Table . Based on the results, the estimated contributions can be reproduced without repeating the whole process (i.e., through simulations). Further the contribution distributions at the family level are evaluated to build a better argument for comparison, since the community ratings also consider a group family contribution.

Table 5: Individual contribution distribution parameters for NPS wave 5

	location	shape	Scale	scale
Beta	0.693809	559.894	-1.2×10^{-24}	2316409
Beta	0.637473	163.3372	-1.4×10^{-25}	475851.1
Pareto	1.851522	17002.53		44731.99

The rates in

Table **2**, are used to simultaneously compute the contribution under the CR approach for all individuals in the validation data. The output for contributions estimation is further analyzed in the test of the hypothesis, which shows that contribution pricing seems to have a near-similar distribution to empirical pricing; however, CR is on a much broader scale than DP from the empirical spending distribution. Furthermore, they are from the same distribution family, see Figure 6 and 7.

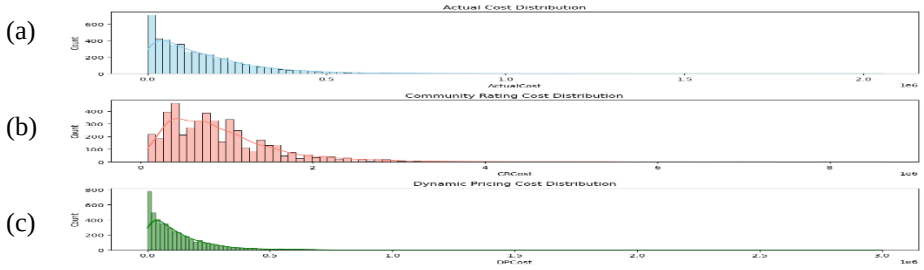

Figure 6: Individual (family level) contribution distribution comparison (a) Actual (b) CR (c) DP

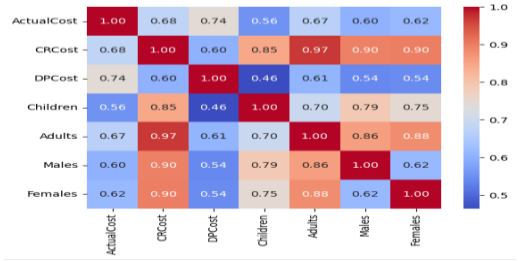

Figure 7: Correlation heatmap

3 Discussion

The results from validation data (NPS wave 5), shows the dynamic pricing pipeline seems to have learned better about the distribution of healthcare costs from previous waves. This is because the estimated contribution rates follow the distribution of empirical healthcare utilization cost, as per Figure 6, which is much better. Data attributes have significant information values for all target variables, and dimensionality reduction could not improve models' performance from the initial feature selection. Furthermore, the evidence from linear models shows a weak linear relation; thus, opting for nonlinear models is appropriate. The RF models have higher computational costs and are prone to overfitting despite producing higher performance than other approaches. The XGB models is adopted, illustrating a better bias-variance tradeoff and significantly lower computational time. Healthcare spending follows very skewed distributions, and extremes exist, which are well learned by a DP model with the implication that high-risk individuals are expected to contribute much more than others; this can limit accessibility to low-income, high-risk individuals. The CR can be a better alternative to distributing extreme costs in this regard. However, this can be normalized to alternative premium principles, such as variance and standard deviation principles, as a significant proportion of the population is expected to make nearly zero contributions. Thus, it can pool the cost that each individual contributes to at least the mean of the projected healthcare cost for the whole scheme. This approach could ensure improved equity in accessing healthcare.

At family level, how various factors could influence individual contributions are analyzed, including gender distribution, age distribution (categorized in two, i.e., adults and children), and cost produced by the two models compared to empirical cost. At baseline, DP cost correlates more to actual cost than the CR approach. This is supported by the distribution of actual values as in Figure 7, whereby CR cost is highly correlated to actual cost, but the distribution of its values is different. The actual cost increases more for adults and females in age and gender categories than for children and males. On the other hand, a similar effect for particular categories is far higher in the CR approach and slightly lower in the DP approach, comprehending the relative effect on accessibility for the two approaches from the actual family use to be different. The test results in Table 6 for hypothesis of Section 1.2.2, against the actual cost for each approach and against one other, portray confident that CR significantly increases the cost from actual healthcare spending while the DP approach reduces it.

Table 6: Hypotheses test summary

μ_0	μ_1	Test statistic	P- value
Actual Cost	Community Rating	-75.3631	0
Actual Cost	Dynamic Pricing	5.519638	3.49×10^{-08}
Community Rating	Dynamic Pricing	76.79101	0

Furthermore, the DP approach also reduces the contribution cost compared to the CR approach. Subsequently, in the dimension of affordability, makes it confident that the

DP approach can significantly provide improved healthcare accessibility compared to CR or OOP approaches.

For sensitivity analysis 10,000,000 simulations are performed to evaluate how the distribution of mean and variance is varied for the OOP, the contributions produced by CR model, and the contributions produced by the DP model. Since the DP approach did not include external expenses incurred outside the actual settlement of the claim, the loading factor is adjusted, as explained in Section 1.2.2 to analyze how the induction of the expenses could affect the results, see Figure 8 and 9.

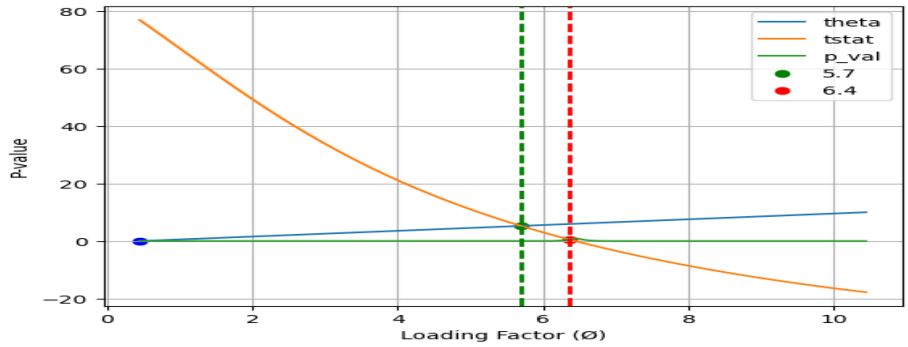

Figure 8: Sensitivity analysis between CR and DP

Figure 8, shows that when the loading factor ranges between 5.66 and 6.35, there is no significant difference between the CR and DP approaches. When it exceeds 6.35, the DP approach significantly increases the healthcare cost thus considered to lower healthcare access.

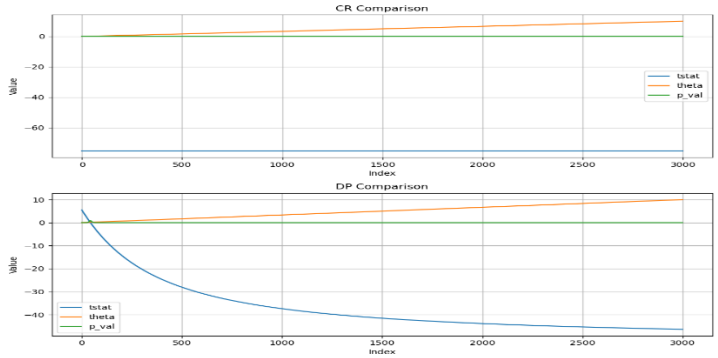

Figure 9: Sensitivity analysis of models to actual cost

In Figure 9, the results show that the loading factor does not impact the CR approach. In the DP approach, when the loading factor is between 0.09 and 0.2, the contribution rates are not significantly different from OOP. At loading above 0.2, the DP rates are significantly higher than OOP, thus, inferring the DP is not appropriate to enhancing healthcare accessibility over OOP.

The results imply that the DP approach provides a promising framework for optimizing healthcare contributions and improving accessibility. By leveraging machine learning techniques and pure premium models, the DP approach demonstrates a nuanced understanding of healthcare cost distributions, effectively balancing individual contributions while ensuring equitable access. The comparison with traditional CR approaches highlights the DP model's superiority in aligning contribution rates with actual costs, potentially reducing individual financial burdens.

Moreover, sensitivity analyses illuminate the DP model's robustness to varying external factors, offering insights into its adaptability and potential limitations. While computational costs and complexities remain challenging, the DP approach presents a viable solution for healthcare financing, warranting further exploration and refinement.

4 Conclusion

A comparative analysis of DP models and CR approach in healthcare financing with a link to healthcare access has been intensively explored through literature review and development of robust regression and classification models for predicting healthcare costs and individual contributions. The statistical techniques (linear regression and logit models) and ML algorithms are used to identify key factors influencing healthcare expenditure. Different tests including KMO, BTS, and log-likelihood ratio test are performed and utilized metrics including MAPE, RMSE, Adjusted R-squared, accuracy score, and F1-scored in data preprocessing and model evaluation, overcoming challenges such as multicollinearity. The analysis, unveiled the differential effect of DP on diverse demographic cohorts, offering insights crucial for shaping equitable healthcare access policies.

The research represents a significant leap toward redefining healthcare financing paradigms, highlighting the transformative potential of DP in fostering accessible and inclusive healthcare systems about CR and OOP. By harnessing the power of machine learning, have not only improved the accuracy of cost prediction but also shed light on the intricate dynamics of healthcare economics.

5 Acknowledgements

The authors are grateful to UDSM for providing supportive research environment and to the NBS in Tanzania for granting permission for use of its data for this research.

References

1. Albawi, S., Alshahrani, L., Albawi, N., Alharbi, R., Alhakamy, A.: Prediction of healthcare insurance costs. In *Computers and Informatics*. Vol. 3, Issue 1 (2023). https://dergipark.org.tr/tr/pub/ci

2. Samiuddin, M., Rajender, G., Varma, S., Kumar, A., Shaik, S. Health Insurance Cost Prediction Using Deep Neural Network. *Asian Journal of Research in Computer Science*, *16*(2), 46–53 (2023). https://doi.org/10.9734/ajrcos/2023/v16i2338
3. WHO (India). Global review of the role of artificial intelligence and machine learning in health-care financing for UHC, (2024).
4. Reid, M., Gupta, R., Roberts, G., Goosby, E., Wesson, P. Achieving Universal Health Coverage (UHC): Dominance analysis across 183 countries highlights importance of strengthening health workforce. *PLoS ONE*, *15*(3) (2020). https://doi.org/10.1371/journal.pone.0229666
5. Darrudi, A., Khoonsari, M., Tajvar, M. Challenges to Achieving Universal Health Coverage Throughout the World: A Systematic Review. In *Journal of Preventive Medicine and Public Health* - Korean Society for Preventive Medicine Vol. 55, Issue 2, pp. 125–133 (2022). https://doi.org/10.3961/jpmph.21.542
6. Neun, S, Sources of Health Care Funding Throughout the Globe. In *ECONOMICS INTERACTIONS WITH OTHER DISCIPLINES*. (2023). https://www.eolss.net/sample-chapters/C13/E6-29-01-03.pdf
7. WHO GHED. *Global Health Expenditure Database* (2024). https://apps.who.int/nha/database
8. Soucat, A., Tandon, A., Gonzales Pier, E. From Universal Health Coverage services packages to budget appropriation: the long journey to implementation. *BMJ Global Health*, *8* (2023). https://doi.org/10.1136/bmjgh-2022-010755
9. Pardo, C., Sabat, J. Equity and Efficiency Effects of Flat Premiums. *SSRN Electronic Journal* (2023). https://doi.org/10.2139/ssrn.4460458
10. Ly, M., Bassoum, O., Faye, A. Universal health insurance in Africa: A narrative review of the literature on institutional models. In *BMJ Global Health* - BMJ Publishing Group.Vol. 7, Issue 4 (2022a). https://doi.org/10.1136/bmjgh-2021-008219
11. Hanafy, M., Mahmoud, O. Predict Health Insurance Cost by using Machine Learning and DNN Regression Models. *International Journal of Innovative Technology and Exploring Engineering*, *10*(3), 137–143 (2021). https://doi.org/10.35940/ijitee.C8364.0110321
12. Kaushik, K., Bhardwaj, A., Dwivedi, A, Singh, R. Article Machine Learning-Based Regression Framework to Predict Health Insurance Premiums. *International Journal of Environmental Research and Public Health*, *19*(13) (2022). https://doi.org/10.3390/ijerph19137898
13. ul Hassan, C., Iqbal, J., Hussain, S., AlSalman, H., Mosleh, M., Sajid Ullah, S. A Computational Intelligence Approach for Predicting Medical Insurance Cost. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/1162553
14. Cu, A., Meister, S., Lefebvre, B., Ridde, V. Assessing healthcare access using the Levesque's conceptual framework– a scoping review. In *International Journal for Equity in Health*-BioMed Central Ltd Vol. 20, Issue 1 (2021). https://doi.org/10.1186/s12939-021-01416-3
15. Ghanad, A. An Overview of Quantitative Research Methods. *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS*. (2023). https://doi.org/10.47191/ijmra/v6-i8-52
16. NBS Tanzania. *National Bureau of Statistics - National Panel Survey* (2024). https://www.nbs.go.tz/index.php/en/census-surveys/poverty-indicators-statistics/national-panel-survey
17. The DHS Program. *The DHS Program – Data* (2024). https://dhsprogram.com/Data/
18. Krejcie, R., Morgan, D. Determining sample size for research activities. 607–610 (1970).

19. Langenberger, B., Schulte, T., Groene, O. The application of machine learning to predict high-cost patients: A performance-comparison of different models using healthcare claims data. *PLoS ONE*, *18*(1) (2023). https://doi.org/10.1371/journal.pone.0279540
20. Benesty, J., Chen, J., Huang, Y., Cohen, I. Pearson correlation coefficient. In *Springer Topics in Signal Processing*- Springer Science and Business Media B.V Vol. 2, pp. 1–4 (2009). https://doi.org/10.1007/978-3-642-00296-0_5
21. McHugh, M. The Chi-square test of independence. *Biochemia Medica*, *23*(2), 143–149 (2013). https://doi.org/10.11613/BM.2013.018
22. Wiley, M., Wiley, J. Univariate Data Visualization. In *Advanced R Statistical Programming and Data Models*-Apress pp. 1–31 (2019). https://doi.org/10.1007/978-1-4842-2872-2_1
23. Pham, T., Pandis, N., White, I. Missing data: Issues, concepts, methods. *Seminars in Orthodontics*, *30*(1), 37–44 (2024). https://doi.org/10.1053/j.sodo.2024.01.007
24. Aguinis, H., Gottfredson, R., Joo, H. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. In *Organizational Research Methods* SAGE Publications Inc. Vol. 16, Issue 2, pp. 270–301 (2013). https://doi.org/10.1177/1094428112470848
25. Kwak, S., Kim, J. Statistical data preparation: Management of missing values and outliers. In *Korean Journal of Anesthesiology* - Korean Society of Anesthesiologists Vol. 70, Issue 4, pp. 407–411. (2017). https://doi.org/10.4097/kjae.2017.70.4.407
26. Singh Rawat, S., Kumar Mishra, A. Review of Methods for Handling Class-Imbalanced in Classification Problems (2023).
27. Van Den Goorbergh, R., Van Smeden, M., Timmerman, D., Ben Van Calster. The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association*, *29*(9), 1525–1534 (2022).. https://doi.org/10.1093/jamia/ocac093
28. Wongvorachan, T., He, S., Bulut, O. A Comparison of Under-sampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information (Switzerland)*, *14*(1) (2023). https://doi.org/10.3390/info14010054
29. Cerda, P., Varoquaux, G. *Encoding high-cardinality string categorical variables* (2019). https://doi.org/10.1109/TKDE.2020.2992529
30. Seger, C. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. In *DEGREE PROJECT TECHNOLOGY* (2018).
31. Shrestha, N. Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4–11 (2021). https://doi.org/10.12691/ajams-9-1-2
32. Taherdoost, H., Sahibuddin, S., Jalaliyoon, N. *Exploratory Factor Analysis; Concepts and Theory* (2020).
33. Orji, U., Ukwandu, E. Machine learning for an explainable cost prediction of medical insurance. *Machine Learning with Applications*, *15*, 100516 (2024). https://doi.org/10.1016/j.mlwa.2023.100516
34. The United Republic of Tanzania. Muswada wa Sheria ya Bima ya Afya kwa Wote wa Mwaka 2022 (Special Bill Supplement No. 8). Dodoma: Government Printer (2022).
35. The United Republic of Tanzania. Sheria ya Bima ya Afya Kwa Wote ya Mwaka 2023. No. 13, 1st December, 2023. ACT SUPPLEMENT To the Gazette of The United Republic of Tanzania No. 48 Vol. 104 Dated 1st December, 2023. Printed By the Government Printer, Dodoma by Order of Government. ISSN 0856 - 035X (2023).
36. Makwetta, H. Viwango kuchangia bima ya afya kwa wote hivi hapa. https://www.mwananchi.co.tz/mw/habari/kitaifa/viwango-kuchangia-bima-ya-afya-kwa-wote-hivi-hapa-3977248. *Mwananchi* (2022).

37. Mori, A. T. Mandatory health insurance for the informal sector in Tanzania—has it worked anywhere! *Frontiers in Health Services*, *3* (2023). https://doi.org/10.3389/frhs.2023.1247301
38. Xu, M., Fralick, D., Zheng, J. Z., Wang, B., Tu, X. M., Feng, C. The differences and similarities between two-sample t-test and paired t-test. *Shanghai Archives of Psychiatry*, *29*(3), 184–188 (2017). https://doi.org/10.11919/j.issn.1002-0829.217070