Modelling Quality of Life among Adults in Gauteng Province, South Africa: A Machine Learning Approach

Muhammad Hoque*

Department of Public Health, Sefako Makgatho Health Sciences University, Pretoria, South Africa

Abstract: *Background:* Quality of Life (QoL) is a complex construct determined by a complex set of health, socio-economic and environmental determinants. Identifying the principal determinants of QoL and determinants that differ between subgroups of the population helps inform population-targeted intervention policies. This work sought to model QoL for adults in the Gauteng Province of South Africa using machine learning methods.

Methods: The cross-sectional analytical study used the QoL Survey (QoL 2023–2024, Round 7) secondary data from the Gauteng City-Region Observatory with more than 14,000 adult resident responses. Five supervised machine algorithms—Logistic Regression, Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN), were trained and tested on accuracy, precision, recall, F1-score, and AUC-ROC.

Results: Logistic Regression performed the best with 81% accuracy followed by SVM and XGBoost. In all the analyses the individual's health, perceived safety, work status appeared to be the most significant QoL predictors using the application of SHAP. Subgroup analysis results indicated model performance variability across levels of schooling, category of dwellings used etc. Prediction accuracy gender-wise differences remained not significant.

Conclusion: Tlevel of satisfaction with the National Government, access to piped water, and toilet type possess the most influential effect on QoL in Gauteng Province. Logistic Regression is still a strong and interpretable QoL predicting model, with the assistance of the support of SHAP-based interpretation. The research suggests the necessity of health-oriented and place-oriented responses in policies to elevate the life quality of the South African cities' populations.

Keywords: Quality of Life, Machine Learning, Logistic Regression, Socio-economic Predictors.

BACKGROUND

Quality of Life (QoL) was conceptualized as a multi-facet construct encompassing physical, mental, social, and environmental well-being [1-3]. Such instruments as the WHOQOL, EuroHIS-QOL-8, and SF-36 are among those diverse methods which are adopted in measuring QoL both in clinical, as well as non-clinical, populations. Conventional regression analyses all verified predictors such as chronic disease, age, education, and income as associated with QoL [4].

Later, machine learning (ML) was adopted in QoL research because ML can identify subtle associations and relationships between sets of high-dimensional variables. As an example, a study that predicted European adult life satisfaction in 50+ using gradient boosting and random forest algorithms, which explained about 53% of variance — an improvement on classical regression models [5]. Another study applied ML to investigate how leisure influences QoL among older Koreans used ML models to predict the happiness index based on several predictive factors [6]. Also, another study predicted SF-36 scores based on biomarkers and sociodemographic variables in Korean older people using elastic net (penalized regression), and their biggest predictors were strength of grip and speed of gait [7].

ML applications in quality measurements for the health in sub-Saharan Africa are limited, but primarily for disease prediction, surveillance, not QoL [8-10]. A review on ML applications in the public health in Africa, advocated balanced, data-based frameworks, while another in their illustration for mental surveillance in Kenya, indicated the ML potential [11,12]. ML modelling for QoL is still underdeveloped, even as demand is rising. Notably, fairness and bias in ML for African health issues are becoming important. Biases were noted in the colonial era and socioeconomic inequalities that could determine model accuracy [13]. Since South Africa's drastic inequalities in socioeconomics are a concern, ML model fairness is particularly important to reduce biases [14,15].

QoL is an important indicator of general well-being and health, covering physical, psychological, social, and environmental aspects. In South Africa, while persistent social and economic developments have occurred, past differences in living conditions, distribution of populations, and access to basic services have continued to adversely affect QoL for the majority of adults. Existing statistical procedures have had poor predictive capacity and have tended not to capture high-order, non-linear associations between many variables that affect QoL. Additionally, few studies have employed more sophisticated ML methods to predict or describe QoL outcomes, at least with recent representative populations of South Africans. The scarcity of solid analytical procedures complicates policymakers and practitioners in their

*Address correspondence to this author at the Department of Public Health, Sefako Makgatho Health Sciences University, Pretoria, South Africa;
E-mail: muhammad.hoque@smu.ac.za

E-ISSN: 1929-6029/26

efforts to identify important determinants and high-risk populations, consequently constraining the generation of properly focused interventions. Accordingly, in response, the present study attempts to engage ML procedures to predict and clarify QoL among adult populations resident in Gauteng Province, based on extensive data from the 2023–2024 QoL Survey. The aim is to construct high-performing ML models capable of revealing the most important socio-demographic, health, as well as environmental predictors, with the final purpose of arriving at evidence-based information for improving well-being in South Africa.

MATERIALS AND METHODS

Study Design and Data Source

This study was designed as an analytical cross-sectional study and utilized secondary data from the QoL Survey 2023–2024 (Round 7), which had been commissioned jointly on behalf of the Gauteng City-Region Observatory (GCRO). The QoL Survey was a biennial household survey aimed at assessing experienced life, subjective well-being, and the socio-economic environment within the Gauteng Province, South Africa. The 2023–2024 round marked the seventh iteration of the survey and consisted of responses from over 14,000 participants, representative at both municipal and ward levels.

The questionnaire provided comprehensive coverage of variables across multiple domains, including income, accommodation, access to facilities, livelihood, safety, health, education, civic engagement, and livelihood satisfaction. These extensive indicators enabled the data set to be used for modelling QoL predictor variables using machine learning algorithms.

Study Population and Inclusion Criteria

The study population included all adult residents of Gauteng Province during the study period when data were collected, aged 18 years and older. Observations that provided complete responses for the dependent variable (self-perceived overall quality of life) and relevant predictor variables were included in the study. Observations with missing values for key variables—particularly the dependent variable—were excluded via listwise deletion or imputation, depending on the observed patterns of missingness.

Outcome Variable

The outcome variable was the self-perceived quality of life (QoL), which respondents rated using a standard 5-point Likert scale ranging from 1 (very poor) to 5 (excellent). To facilitate model performance and interpretability, the outcome was transformed into a binary variable: scores of 1 and 2 were grouped as "Low QoL," while scores of 4 and 5 were grouped as "High QoL." Responses rated as 3 ("Neutral") were excluded from the primary classification models but included in sensitivity analyses.

Independent Variables

Socio-demographic and household-level variables were included as independent variables. Demographic variables included gender, population group, and age. Socio-economic variables included employment status, income, educational attainment, and household asset ownership. Environmental factors considered were access to safe water, electricity, housing structure, and transport. Health-related variables included self-rated health, access to medical care, and presence of chronic disease. Social variables such as perceptions of safety, social support, and civic engagement were also considered. Geographic identifiers, such as municipality and ward-level position, were also included.

Data Preprocessing

Before model development, the dataset underwent several preprocessing steps to ensure quality and integrity. The initial step involved removing duplicate rows and identifying outliers, followed by appropriate treatment for missing data. Continuous variables with missing values were imputed using Multiple Imputation by Chained Equations (MICE), while categorical variables were imputed using mode values.

Categorical variables were encoded using one-hot or ordinal encoding, depending on their nature, and continuous variables were normalized for uniform scaling. The final dataset was randomly split into training and testing sets using a 70/30 ratio, and class-balancing techniques were applied to address class imbalance in the outcome variable.

Machine Learning Models

Five supervised classification algorithms were developed and compared in this study. These included Logistic Regression (as the baseline model), Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN). Optimal hyperparameters for each model were determined through grid search with 5-fold cross-validation on the training set.

Model Evaluation

Model performance was assessed on the test set using multiple evaluation metrics to provide a comprehensive understanding of each model's strengths and weaknesses. These metrics included accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic curve (AUC-ROC). Calibration of the models was assessed through reliability curves and Brier scores. The best-performing model was further interpreted using SHAP (SHapley Additive exPlanations) values, enabling identification of individual feature contributions to the model's predictions.

Accuracy of subgroups was assessed based on the comparison between predicted and true values of QoL results for the specific subgroups in the test set. One single model was applied to all the subgroups. There is high variability in subgroup accuracy because the data is different from subgroup to subgroup.

Post-Hoc Model Interpretation via SHAP

To facilitate model interpretability, the use of SHapley Additive exPlanations (SHAP) technique has been considered as a post-hoc approach that would provide measures of how much each feature contributes to the model predictions. In particular, SHAP technique breaks down predictions into additive component contributions as compared to some base-line value. Specific SHAP techniques have been used for specific models – namely, TreeSHAP for tree-based models (Random Forest and XGBoost) and Deep SHAP for Artificial Neural Network. Logistic Regression and Support Vector Machine require the use of Kernel SHAP.

Ethical Considerations

The research was conducted using anonymized, open-access secondary data. Ethical approval had been obtained before the commencement of the analysis from the Ethics Committee at the Sefako Makgatho Health Sciences University (SMUREC) (Ref No: SMUREC/H/399/2025:IR). All procedures complied with ethical standards and ensured data confidentiality. Data handling adhered strictly to the South African Protection of Personal Information Act (POPIA) to uphold ethical integrity.

RESULTS

More than half of the respondents were female (51.3%), some respondents had no source of income or preferred not to answer (30.8%), and most were coloured (78.5%). Most of the respondents lived in a dwelling type made of brick or concrete (67.8%), more than half (52.0%) do not engage in any type of work, business or activity and almost half (45.7%) were dissatisfied with money. Above half (61.0%) of respondents always have clean water and only a few (7.1%) reported that the loadshedding had no impact at all on their study (Table **1**).

According to Table **2**, all models demonstrate relatively similar moderate performance in QoL prediction, with their accuracy lying between 0.75 to 0.81. However, the best classification accuracy (0.81) was reached by the Logistic Regression model, which is characterized by the balanced combination of precision, recall, and F1-score. The next two models demonstrating high classification performance were SVM and XGBoost. Their classification results are characterized by the balanced class-specific metrics (precision, recall, and F1 ≈ 0.80) and therefore represent stable and unbiased predictions.

A Random Forest model demonstrates relatively high but slightly worse performance compared with the above-mentioned models. In turn, the ANN is associated with the lowest classification performance, demonstrated by its lowest accuracy (0.75) and weak F1-scores.

It should be noted that for all models, there is a relatively high similarity between class-specific metrics, namely, precision and recall, between low (0) and high (1) QoL. As stated above, this is an important factor for binary classification, during which both type I and II errors need to be estimated.

Even though RMSE and R^2 have been calculated, it is not recommended to use them since they are the indicators of regression and may be inappropriate for classification tasks. Namely, RMSE represents the prediction error in probabilities, whereas R^2 is associated with the variance explained. Moreover, such metrics are not reliable for binary classifications.

Low R^2 scores (≤ 0.23) demonstrate that even though classification results of all considered models are acceptable, the models explain a small part of the total variability of QoL indicators.

Prediction capacities of each of the model- SHAP analysis was used to determine the relative importance of selected predictors. The figure **1** shows the factors that are most strongly related to the QoL. The top predictors are the level of satisfaction with the National Government, access to piped water, and toilet type. This suggest that these are strong predictors of the QoL.

Some important predictors such as access to piped water, toilet type, level of satisfaction with the National Government, type of healthcare, access to clean water are shown. The model predicts an individual's QoL based on an individual's access to basic infrastructure

Table 1: Socio-Demographic Characteristics of the Population

Demographic variable	№	%
Sex		
Male	6,713	48.7
Female		
Income	7,082	51.3
R1-R800	332	2.41
R801-R3 200	2,968	21.51
R3 210- R12 800	3,851	27.91
R12 801- R25 600	1,280	9.28
R25 601- R51 200	702	5.09
R51 201 and more	419	3.04
No income/Don't know/Prefer not to answer	4,243	30.76
Population group	10,834	78.54
Black African	584	4.23
Coloured*	379	2.75
Indian/Asian	1,978	14.34
White	20	0.14
Other		
Dwelling type of household		
House, brick, or concrete structure	9,357	67.83
Traditional dwelling, hut, or structure	51	0.37
Flat or apartment in a block of flats	1,169	8.47
Cluster house in a complex	175	1.27
Townhouse (semi-detached house in a complex)	173	1.25
Semi-detached house not in a complex	58	0.42
House, flat or room separate from main	630	4.57
Informal dwelling or shack in backyard	753	5.46
Informal dwelling NOT in backyard	1,114	8.08
Room or flat which is part of main dwelling	172	1.25
Caravan or tent	3	0.02
Unit in a retirement home or barracks	21	0.15
Hostel	119	0.86
Major assets such as cars and furniture insured		
All are	967	7.01
Most are	1,005	7.29
Some are	2,756	19.98
None are	9,067	65.73
Do you do any type of work, business, or activity.		
No	7,179	52.04
Yes	6,616	47.96
How satisfied are you with money?		
Very satisfied	283	2.05
Satisfied	2,172	15.74
Neither satisfied nor dissatisfied	943	6.84
Dissatisfied	6,311	45.75
Very dissatisfied	4,086	29.62
How is this dwelling owned?		
Owned, but paying off a bond	1,789	20.69
Owned, fully paid off	3,098	35.83
Free RDP house	1,844	21.33
Built on land you own using own or family	1,031	11.92
Built using government housing subsidy	172	1.99
Transfer of title deed of existing government	152	1.76
Inherited	421	4.87
Other	139	1.61

(Table 2). Continued.

Demographic variable	№	%
Is the water you receive clean?		
Always	8,416	61.01
Usually	2,997	21.73
Sometimes	2,133	15.46
Hardly Ever	163	1.18
Never	86	0.62
How did the loadshedding impact your study?		
Significant impact	4,406	31.94
Moderate impact	2,000	14.50
Very limited impact	977	7.08
No impact at all	1,605	11.63
Not applicable, no scholars or students	4,807	34.85

*People of mixed ethnic origin.

Table 2: Performance of Machine Learning Models in Predicting QoL Outcomes

Model	Precision (Low)	Precision (High)	Recall (Low)	Recall (High)	F1-score (Low)	F1-score (High)	Accuracy	RMSE*	R²*	Support (Low)	Support (High)
Support Vector Machine	0.80	0.79	0.79	0.80	0.80	0.80	0.80	0.45	0.18	2070	2069
Logistic Regression	0.80	0.81	0.79	0.83	0.79	0.82	0.81	0.49	0.23	1300	1459
Random Forest	0.79	0.79	0.79	0.78	0.79	0.79	0.79	0.46	0.14	2070	2069
Artificial Neural Network	0.74	0.75	0.75	0.74	0.75	0.74	0.75	0.50	-0.02	2070	2069
Extreme Gradient Boosting	0.80	0.81	0.81	0.80	0.81	0.80	0.80	0.44	0.21	1380	1379

*0= QoL (Low), 1= QoL (High).

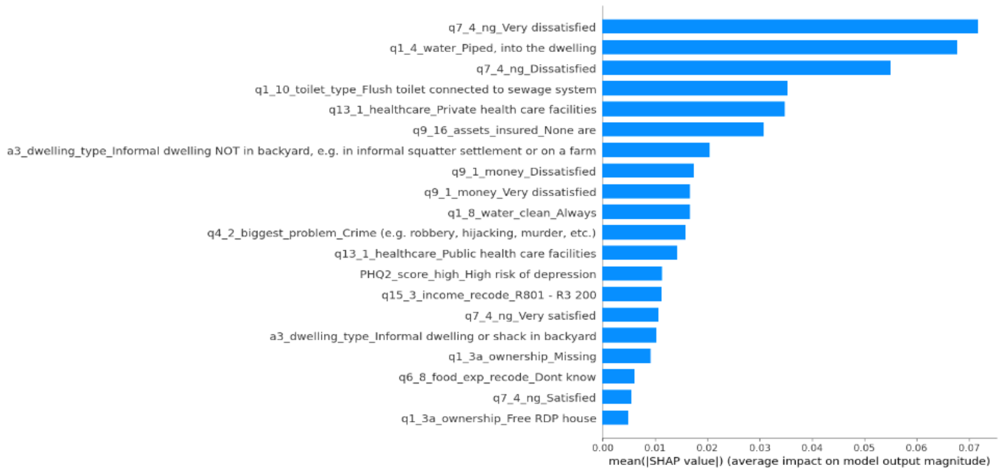

Figure 1: Logistic Regression Model.

like clean water, sanitation, satisfaction with the government and financial security to make its predictions (Figure **2**).

Figure **3** shows the relationship between the top predictors of QoL- access to piped water, access to different health care facilities, having a flush toilet connected to a sewage system and having no assets insured.

The most important factor is the level of satisfaction with the National Government. Other important factors

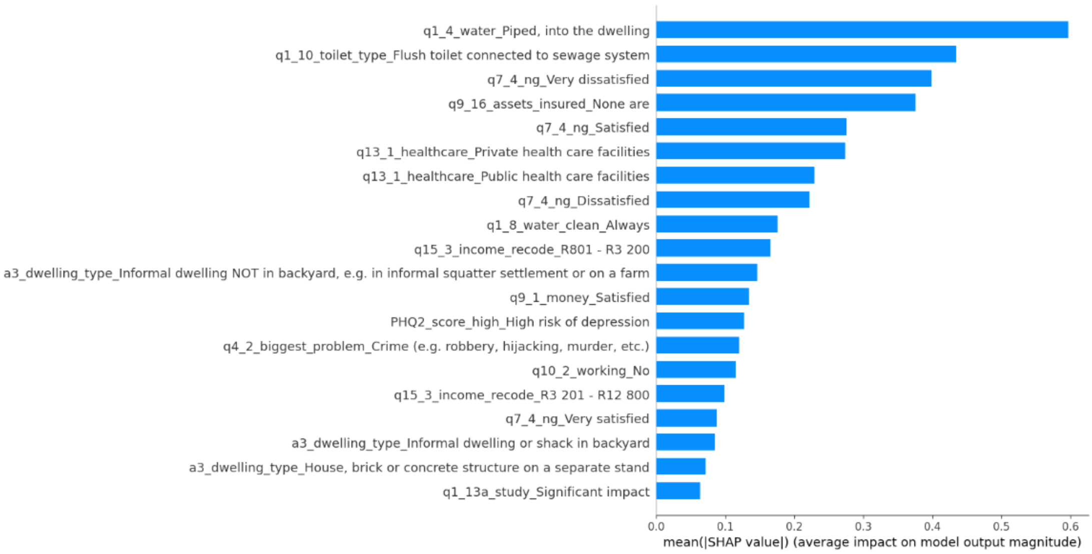

Figure 2: Extreme Gradient Boost Model- SHAP Feature Importance Plot.

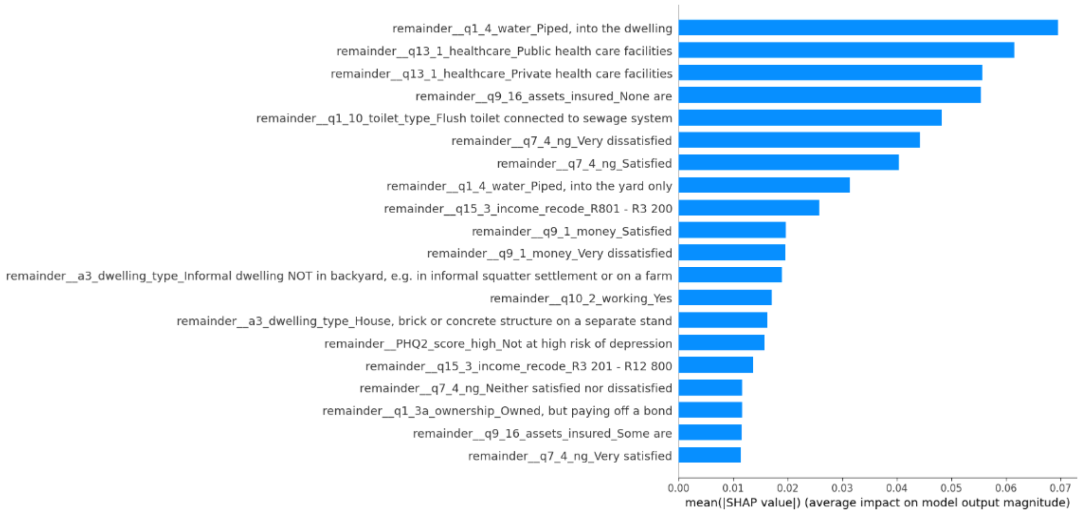

Figure 3: Random Forest Model- SHAP Feature Importance Plot.

include access to piped water, not having insured assets, healthcare access- private healthcare and access to a flush toilet which signifies sanitation. These factors have the biggest impact on the model (Figure **4**).

The important predictors are access to piped water, access to healthcare facilities, level of satisfaction with the National Government's policies. Others are satisfaction to money, income. Overall, the model depends to an individual's access to basic infrastructure and satisfaction with the government and living conditions to make predictions (Figure **5**).

Using the Logistic Regression Model to further predict how some selected variables perform as predictors for QoL through SHAP analysis. It was observed again that the most important predictors of an individual's QoL have to do with their access to basic infrastructure and their level of satisfaction as regards the government. The two most important features have to do with the level of satisfaction of the activities of the government, this is followed by access to toilet type, access to piped water and access to healthcare facilities. Overall, predictions are primarily influenced by an individual's access to basic services and level of satisfaction about the government's activities (Figure **6**).

Figure **7** shows the level of importance of different factors in influencing the model's predictions. The model is most accurate on groups that are at the top of the chart which are 'Informal dwelling not in the backyard' and income within the R51 201 bracket, while the lowest accuracy was observed for subgroups like 'unit in a retirement home or barracks' and 'semi-detached house not in a complex' (Figure **7**).

In this subgroup analysis, there is a consistent disparity in the model's performance across different levels of sanitation and water infrastructure. The model

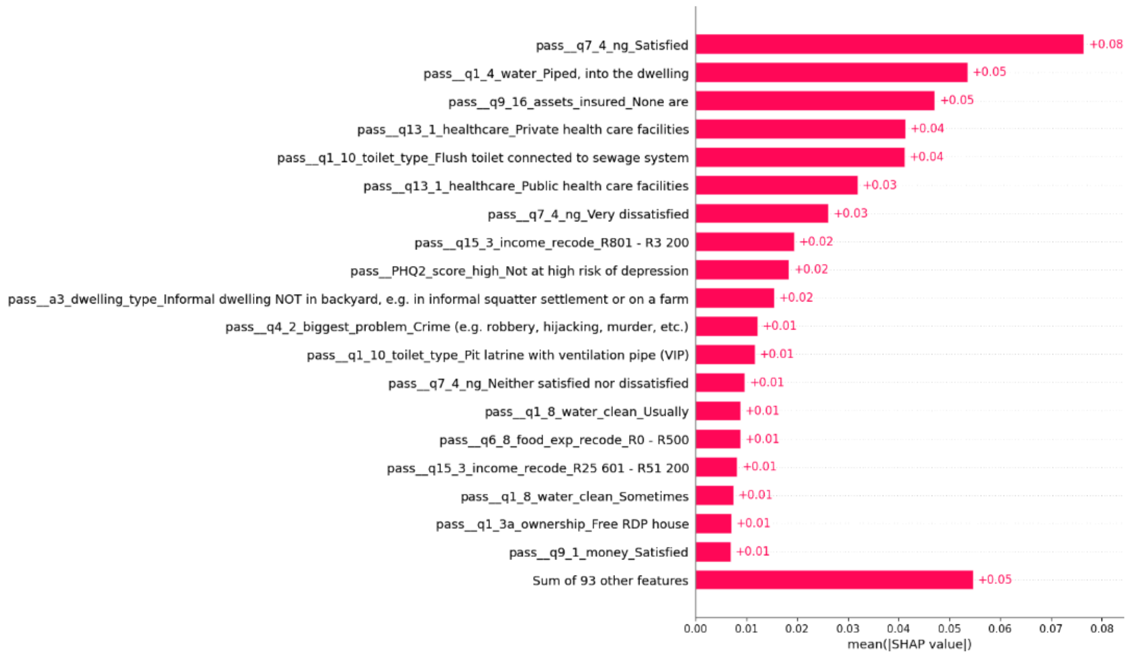

Figure 4: Support Vector Model- SHAP Feature Importance Plot.

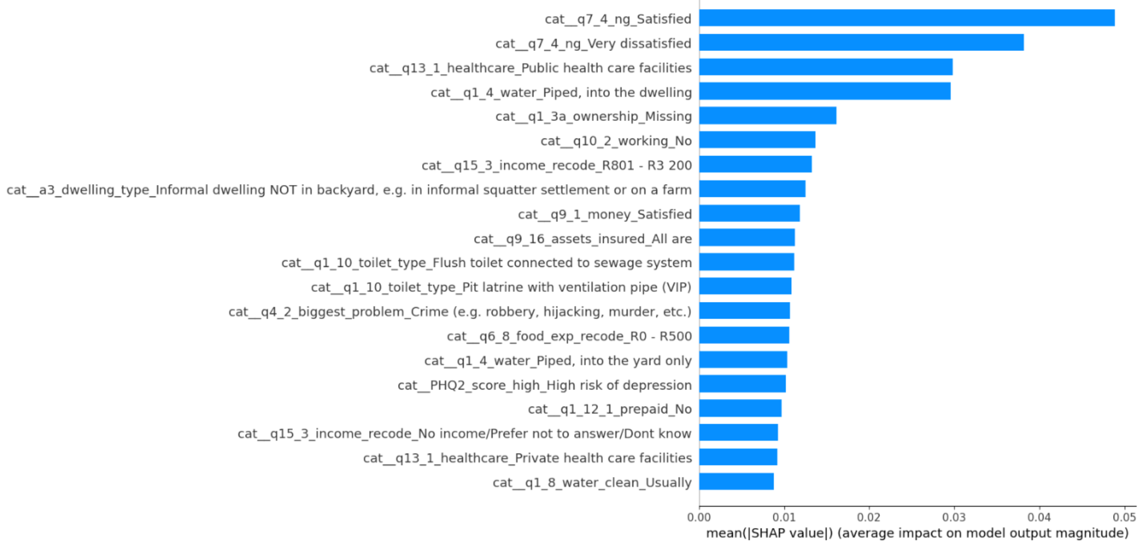

Figure 5: Artificial Neural Network Model- SHAP Feature Importance Plot.

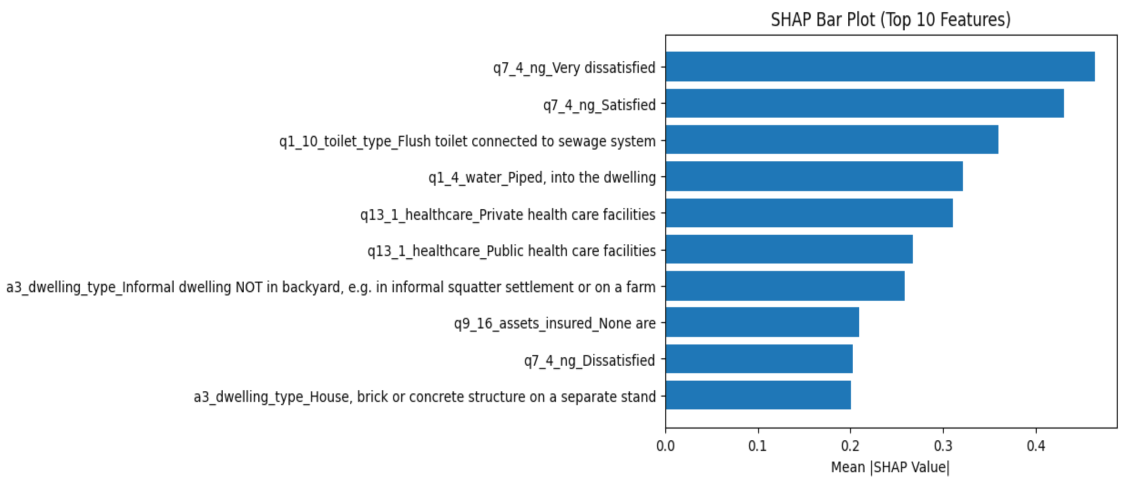

Figure 6: Performance of some selected variables as predictors for the Quality of Life.

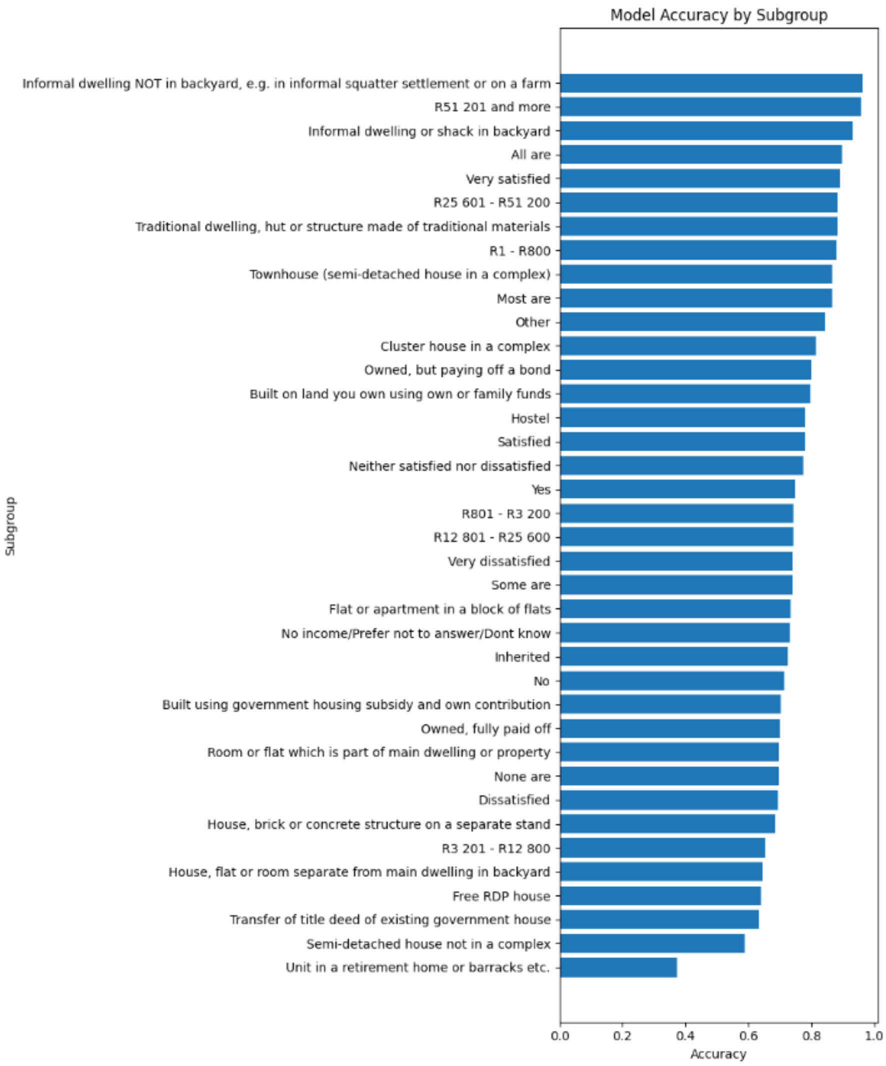

Figure 7: Sub-group analysis of dwelling type, assets insured, whether one's assets are insured and income.

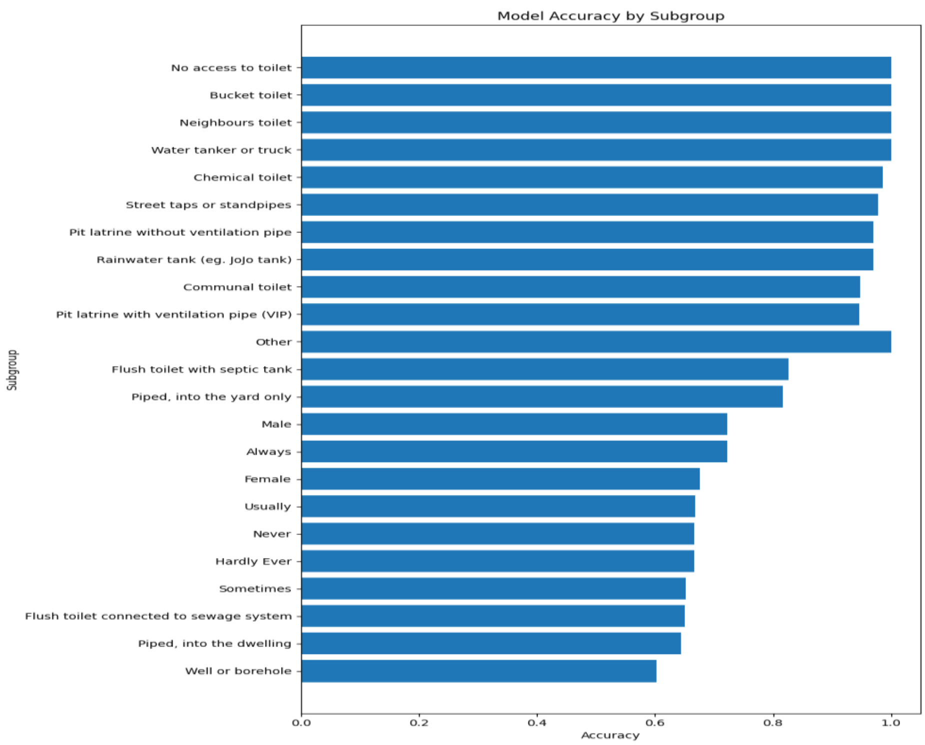

Figure 8: Sub-group analysis of gender, access to clean water, source of water and toilet type.

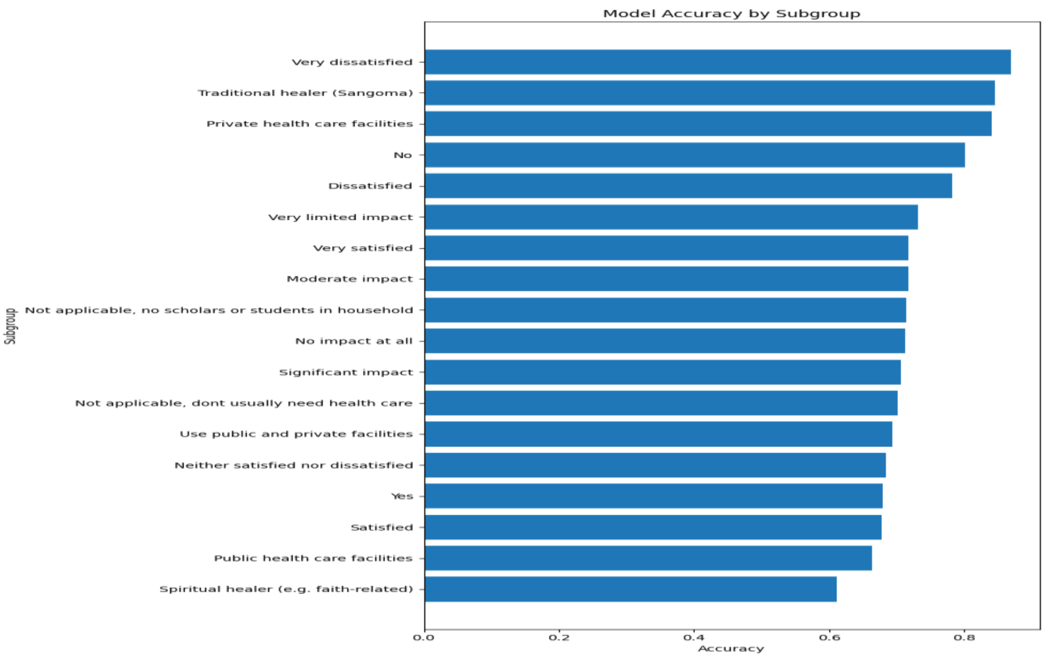

Figure 9: Sub-group analysis of study type, satisfaction with dwelling place, how the challenges of load shedding affected their study and choice of healthcare facility.

performs with the highest accuracy on subgroups that have less developed sanitation- 'No access to toilet', 'bucket toilet' and 'neighbours' toilet' which suggests that the model is good at identifying patterns in this context (Figure **8**).

The model has the highest accuracy for data related to level of satisfaction to the National Government, impact of load shedding on one's study and access to health care facilities, particularly private (Figure **9**).

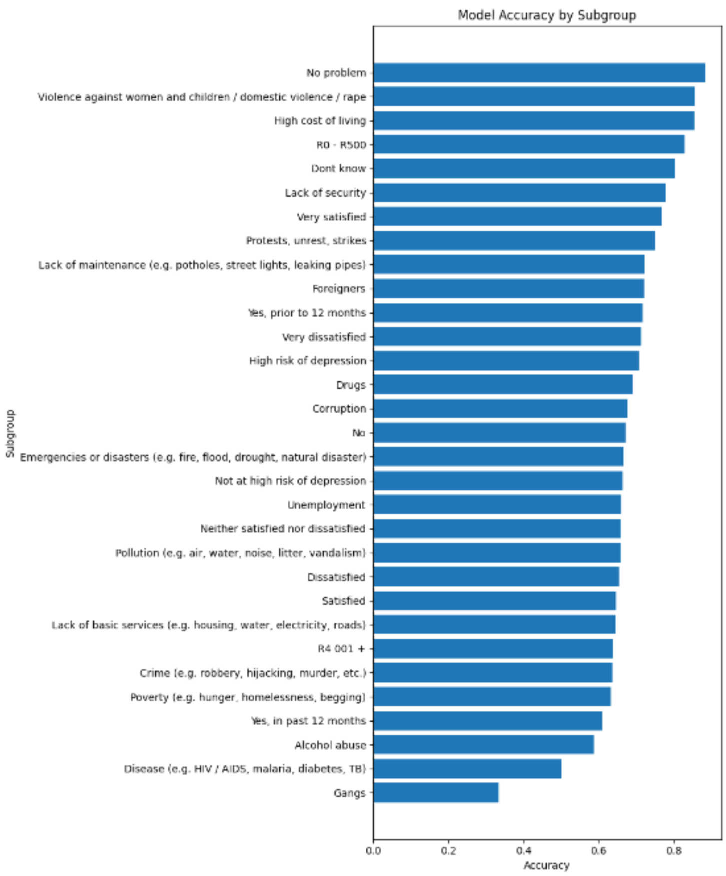

Figure 10: Sub-group analysis of satisfaction with the performance of the National Government, biggest problem facing the neighbourhood, and PHQ score indicative of high risk of depression.

The model's strength in predicting other outcomes is shown in Figure **10**. It is shown that factors including violence against women, lack of security, satisfaction with the National Government has the highest accuracy (Figure **10**).

Overall, the subgroup analysis reveals the strength and weakness of the Logistic Regression model across different categories. The model has high accuracy for the top features, and this means the model's predictions are highly reliable for these features and strong patterns were found in the data related to it. On the other hand, the features at the bottom imply that the model's predictions for these features are not as reliable. For instance, more data can be collected for features in this category to perform more investigation.

DISCUSSION

These results point out that out of the five ML models used—Logistic Regression, Random Forest, SVM, XGBoost, and ANN. Logistic Regression emerged tops in predicting QoL outcomes with 81% accuracy. This is quite interesting considering the fact that ensemble techniques such as XGBoost were thought to outperform linear models. What is most shocking is the fact that XGBoost only reached 67% accuracy, and SVM and Random Forest only reached 69% and 65% respectively. ANN performed the worst with 44% accuracy, something that arguably persists due to its highly sensitive nature to the choice of hyperparameters and higher requirement on the data to generalize well.

These findings echo prior work uncovering that even with the richness of sophisticated algorithms in machine learning, Logistic Regression continues to be a contender and interpretable answer to health-related prediction tasks. An extensive meta-analysis was performed and asserted that clinical prediction models from tabular structured data do not necessarily benefit from the application of logistic regression [16]. The stability and the interpretation of the logistic regression make it a favoured instrument whenever the task involves not only the prediction but also the identification of the influence of particular predictors on outcomes. In addition to that, the relatively modest performance of ANN in this test should be attributed to the relatively modest size of the dataset or the inadequacy of suitable training iterations, a common failure in the application of neural networks involving large datasets for the efficient identification of nonlinear behaviours [17].

Health and safety were found among all the machine learning models applied. The finding provides support for international and South African research on the overriding influence of the state of health in the determination of perceived well-being [18,19]. The respondents who chose good health chose the higher scales of life satisfaction because of the lowered body limitations, lowered psychological distress, and the higher social and economic life involvement [5]. Safety, especially the experience of environment and individual safety, equally determines QoL. In conditions in which the safety of the public is compromised or questionable, the population displays higher levels of stress and limited mobility, and this impacts the subjective well-being negatively [20].

Similar to other models, and especially the Random Forest and the XGBoost model, safety and health appeared as the most influential predictors and also produced interaction effects, revealing the combined effect on QoL to be reinforced or penalized whenever the two variables coexist poorly or positively. The interaction comes out well with systems thinking on health and social well-being, such that the interaction of determinants presents nonlinear effects on outcome [21]. The interaction plots from the SHAP confirmed the same opinion through the illustration of the coexistence of good health and higher safety perception, reinforcing QoL predictions strongly.

Apart from health and safety issues, socio-economic factors like jobs, incomes, and residential satisfaction also rightly influence QoL. Workers also tended to have higher QoL, perhaps owing to higher financial security and social integration. This aligns with Sen's notion of development as an important concept of freedom in which material and social conditions could improve an individual's opportunities [22]. The income levels also registered a highly positive correlation with QoL, as access to health care and other life-improving services is routed through financial channels. Home ownership and residential satisfaction also proved to be influential determiners. The results support earlier studies who identified material standards of living and impressions regarding the adequacy of houses as the dominant components of subjective well-being [23,24].

Notably, from the results of the XGBoost SHAP was the value attributed to access to prepaid electricity meters and coverage from insurance. Those are surrogates for financial prudence and planning. Insured or homeowners who enjoy reliable access to the most important utilities most likely enjoy higher control over life and less everyday stress, and thus higher QoL scores. Educational levels and particularly postgraduate levels also correlated with higher accuracy in the subgroup analyses and indicate that higher-educated persons probably enjoy more disciplined life practices or life self-awareness and thus more stable QoL scores. The finding corroborates reports, revealing the correlation between the schooling level and life satisfaction and health literacy [25].

Subgroup analysis also revealed substantial heterogeneity in model performance between population subgroups. For instance, accuracy differed between the unemployed and the employed and between the two groups, the employed had slightly higher rates of prediction. This difference may be caused by the higher socio-economic consistency and routine of the employed population, whose QoL would accordingly be relatively easy to model [2]. Differences also prevailed between the model's performance in the district municipalities and reflecting spatial inequalities in determinants of QoL and possible differential distribution of resources and infrastructure. Spatial inequities of the aforementioned kind have been documented in South African urban research on the basis of differential service delivery and access to opportunity that differ noticeably between municipal boundaries [26].

CONCLUSION

Overall, the findings show that health, safety, financial security, and environmental sufficiency comprise the foundation of QoL for adult populations in the Gauteng Province. The inclusion of the use of the application of the SHAP analysis in the present work filled the gap between algorithmic forecasting and the use of policies and facilitated the explanation and interpretation of the intricate ML models. The evidence generated in the present work provides a sturdy empirical basis for policies for public health, public safety in the cities, and policies on employment to focus on for the betterment of the overall quality of life.

DECLARATIONS ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval was obtained before the commencement of the analysis from the Ethics Committee at the Sefako Makgatho Health Sciences University (SMUREC).

CONSENT FOR PUBLICATION

Not Applicable. The research was conducted using anonymized, open-access secondary data.

DATA AVAILABILITY STATEMENT

Data is publicly available.

CONFLICT OF INTERESTS

The authors have no competing interests.

FINANCIAL SUPPORT

Not Applicable.

CONTRIBUTIONS

MH conducted the literature review, abstract writing, discussion and manuscript writing.

ACKNOWLEDGEMENTS

I would like to thank Tosin Omoleye for analysing the data.

REFERENCES

[1] Wang X, Li Y. Correlates of quality of life, happiness and life satisfaction among European adults older than 50 years: A machine-learning approach. Psychogeriatrics 2023; 22: 523-31.

[2] Graham C, Nikolova M. Bentham or Aristotle in the development process? An empirical investigation of capabilities and subjective well-being. World Dev 2015; 68: 163-79. https://doi.org/10.1016/j.worlddev.2014.11.018

[3] Veenhoven R. The four qualities of life: ordering concepts and measures of the good life. J Happiness Stud 2000; 1(1): 1-39. https://doi.org/10.1023/A:1010072010360

[4] Pinquart M, Sörensen S. Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. Psychol Aging 2000; 15(2): 187-224. https://doi.org/10.1037/0882-7974.15.2.187

[5] Prati G. Correlates of quality of life, happiness and life satisfaction among European adults older than 50 years: a machine-learning approach. Arch Gerontol Geriatr 2022; 103: 104781. https://doi.org/10.1016/j.archger.2022.104791

[6] Kim EJ, Kang HW, Park SM. Leisure and happiness of the elderly: a machine learning approach. Sustainability 2024; 16(7): 3229. https://doi.org/10.3390/su16072730

[7] Lee SH, Choi I, Ahn WY, Shin E, Cho SI, Kim S, *et al*. Estimating quality of life with biomarkers among older Korean adults: a machine-learning approach. Arch Gerontol Geriatr 2020; 87: 103992. https://doi.org/10.1016/j.archger.2019.103966

[8] Gómez-Olivé FX, Thorogood M, Clark BD, Kahn K, Tollman SM. Assessing health and well-being among older people in rural South Africa. Glob Health Action 2010; 3: 2126. https://doi.org/10.3402/gha.v3i0.2126

[9] Ralston M, Schatz E, Menken J, Gómez-Olivé FX, Tollman S. Policy shift: South Africa's old age pensions' influence on perceived quality of life. J Aging Soc Policy 2019; 31(2): 138-54. https://doi.org/10.1080/08959420.2018.1542243

[10] Van Biljon L, Nel P, Roos V. A partial validation of the WHOQOL-OLD in a sample of older people in South Africa. Glob Health Action 2015; 8: 25820. https://doi.org/10.3402/gha.v8.28209

[11] Kabiru G, Zainab S, Kingsley G, Abubakar W, Zayyanu S, Adam BA, *et al*. Harnessing machine learning for predictive healthcare: a path to efficient health systems in Africa. Health Inform Inf Manag 2025; 1(1): 1-10. https://doi.org/10.17352/hiim.000001

[12] Tshimula JM, Kalengayi M, Makenga D, Lilonge D, Asumani M. Artificial intelligence for public health surveillance in Africa: applications and opportunities. arXiv 2024. https://arxiv.org/abs/2402.12345

[13] Asiedu M, Dieng A, Haykel I, Rostamzadeh N, Pfohl S, Nagpal C, *et al*. The case for globalizing fairness: a mixed methods study on colonialism, AI, and health in Africa. arXiv 2024. https://doi.org/10.1145/3689904.3694708

[14] World Bank. South Africa economic update: towards inclusive growth. Washington, DC: World Bank 2021. Available from: https://www.worldbank.org/en/country/southafrica/report/economic-update2021

[15] Asiedu MN, Dieng A, Oppong A, Nagawa M, Koyejo S, Heller K. Globalizing fairness attributes in machine learning: a case study on health in Africa. Fairness in Global Health, 2023; 1-14. https://doi.org/10.1145/3689904.3694708

[16] Christodoulou E, Ma J, Collins G, Steyerberg E, Verbakel J, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019; 110: 12-22. https://doi.org/10.1016/j.jclinepi.2019.02.004

[17] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019; 380(14): 1347-58. https://doi.org/10.1056/NEJMra1814259

[18] The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL). Soc Sci Med 1995; 41(10): 1403-9. https://doi.org/10.1016/0277-9536(95)00112-K

[19] Ramlagan S, Peltzer K, Phaswana-Mafuya N. Quality of life and associated factors among older adults in South Africa. Glob Health Action 2021; 14(1): 1868700.

[20] Dolan P, Metcalfe R. The relationship between innovation and subjective wellbeing. Res Policy 2012; 41(8): 1489-98. https://doi.org/10.1016/j.respol.2012.04.001

[21] Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health 2011; 101(9): 1627-34. https://doi.org/10.2105/AJPH.2011.300149

[22] Sen A. Development as freedom. Oxford: Oxford University Press; 1999.

[23] Cummins RA. Subjective wellbeing as a social indicator. Soc Indic Res 2018; 135: 879-91. https://doi.org/10.1007/s11205-016-1496-x

[24] Gough KV, Langevang T, Namatovu R. Researching entrepreneurship in low-income settlements: the strengths and challenges of participatory methods. Environ Urban 2014; 26(1): 297-311. https://doi.org/10.1177/0956247813512250

[25] Organisation for Economic Co-operation and Development (OECD). How's life? Measuring well-being. Paris: OECD Publishing 2020.

[26] Roberts B, Struwig J, Gordon S. Quality of life and public services in South Africa: the public perception gap. HSRC Rev 2021; 19(1): 6-9.

Received on 10-06-2026 Accepted on 14-07-2026 Published on 12-08-2026

https://doi.org/10.6000/1929-6029.2026.15.32