GSJ: Volume 11, Issue 12, December 2023, Online: ISSN 2320-9186

www.globalscientificjournal.com

Features that affect customers' decision to do transactions in the near future in mobile operators

Mgeni Athuman, Dr. Joseph Mushi, Dr. Jimmy Mbelwa

Abstract

The challenge of retaining customers and preventing churn has become increasingly complex for mobile operators, given its potential impact on both revenue streams and long-term corporate reputation. Amidst growing concerns about data confidentiality, organizations are more cautious than ever in sharing customer details. This study addresses this pressing issue by identifying key features influencing customers' decisions to transact with mobile operators in the near future. A classification model is constructed based on these identified features, leveraging machine learning predictive models, including Decision Tree, Random Forest, Logistic Regression, and AdaBoost.

The research reveals that AdaBoost outperformed other models, achieving an accuracy of 65.0%. Primary data collection involved a questionnaire administered to 500 students from selected universities. The analysis highlights DO_WHAT_THEY_SAY, GET_THROUGH, HIGH_SWITCHING_ENERGY_TIME and FREE_COMPLAINTS *as prominent features determining a customer's inclination to engage in future transactions. This paper significantly contributes to existing literature by incorporating customer behavior insights into mobile operator transactions, expanding knowledge on customer retention and its crucial determinants.*

Building on these behavioral findings, the study recommends that operators in Tanzania proactively adjust their strategies, making early efforts to retain customers showing signs of falling below a threshold value. This proactive approach could potentially mitigate the decline in transaction activities before customers decide to reduce their engagement with the operator. Early detection of the major features influencing customer retention is crucial to implementing effective preventive measures.

Keywords: *Customer behaviour; Customer churn; Customer retention; Mobile operator.*

1.0 INTRODUCTION

Customer churn (CC) refers to loss of customers from an organization over a given time period (Amin *et al*., 2019; Kaur, 2017; Ullah *et al*., 2019; Umayaparvathi & Iyakutti, 2012). Such loss leads to lost customers abandoning the services and products offered which directly decreases revenue of an operator in the future.

Two categories of CC exist namely involuntary and voluntary. Involuntary churners are those customers that are purposely removed by a company itself due to fraud, inactive, zero usage of services, or non-payment of utilized services among others while voluntary churners are those customers that decide on their own to terminate their service with a respective business (Kaur, 2017; Shaaban *et al*., 2012; Townsend & Nilakanta, 2019). Voluntary churn is further divided into deliberate churn and incidental churn. While deliberate churn is mainly a result of economic factors such as price sensitivity, technological factors, unsatisfactory customer service factors and other inconvenience related factors, incidental churn is caused by unexpected changes in customers' lifestyle such as decreased financial status, geographical changes, and the likes (Townsend & Nilakanta, 2019). With both deliberate and incidental, customers may silently leave the company or reduce their transaction usage thus causing significant losses in revenue as well as reputation.

In 2020, CC rate in the Western telecom industries was in the mid 25% costing organizations $10 billion annually (O'Dea, 2020). Also, in Tanzania, Smart Mobile Phone Company was brought to halt in the late 2019 as a result of CC (Lancaster, 2020).

Customer retention (CR) refers to the activities that an organization undertakes in order to reduce CC and retain their profitability (Ahn *et al*., 2006; Alshurideh, 2016; ; Rather, 2019). It has been largely accepted that active customers once served well, not only are easier and less expensive to retain and do business with them compared to new acquired customers, but are also likely to spread positive words of mouth (PWoM) and the probability of selling to such active customers is at least 40% more likely than to selling to a new customer (Amatare & Ojo,2020.; Chen-Hung *et al*., 2017; Oblander *et al*., 2020; Rather, 2019)

However, regardless of efforts made to improve services and customer interactions to improve retention over a given period of time, customers will tend to reduce transactions in a given business mainly due to reasons that may be related to voluntary or involuntary churn. Thus to be competitive, organizations should constantly improve relationships that will improve retention.

Based on consumer behaviour theory, a customer may decide to purchase or not to purchase a particular service or commodity. Consumer behaviour has been explained as the process where a customer decides what, when, where, how, and from whom to purchase products or service (Chopra *et al*., 2020). The reason for CR or CC in mobile operators may be subjected to customer behaviour (Chopra *et al*., 2020).

However, customer behaviour is a product of several other factors, notably among them is the quality of services offered to them by an organization. Thus in order to improve the chances of retaining customers and make them see more value in the services provided, quality of services offered to them should be of high priority.

Several techniques have been deployed to satisfy customers and retain them. For instance, in an effort to retain and attract new customers, mobile phone companies in Tanzania have been employing techniques such as prize money, cost reduction during certain selected hours, provision of free air time, and cooperate social responsibilities that aim at building companies image in the society. Despite the efforts, no proofs have been shown to improve CR resulting from such techniques introduced (TCRA, 2019; TCRA,2020*)*.

Hence, most mobile operators relying solely on such traditional methods to retain customers are not much useful. Understanding a customer's behavioural path and the factors leading to CR or CC using machine learning models would provide a better chance of minimizing CC and improving CR. This study intended to determine features would lead to customer retention in mobile operators in the near future.

Problem Statement

Despite ongoing efforts, traditional methods employed by most mobile operators to retain customers have shown limited effectiveness. Understanding the customer's behavioral path and the factors leading to CR or CC using machine learning models could provide a more effective strategy for minimizing CC and improving CR.

Objectives of the study

The Features that affect customers' decision to do transactions in the near future in mobile operators possess the following objectives:

1. **To Identify Features for Customer Retention:**
    - To determine the features that lead to customer retention in mobile operators in the near future.
2. **To Apply Machine Learning Models:**
    - To utilize machine learning models to analyze customer behavior and predict features influencing customer retention.

Justification

Traditional customer retention methods have shown limitations, necessitating exploration of more advanced techniques such as machine learning models. Understanding the features influencing customer retention through these models can provide a data-driven approach, potentially leading to more successful customer retention strategies in mobile operators.

2.0 RELATED WORK

Rahman & Chowdhury, 2022 did a study to analyze factors of CR for young customers and found that price perception, customer satisfaction, switching barriers, and brand image were the strongest CR determinants.

Saputro *et al*., 2021 used three (3) classification methodologies, Fast Large-Margin, Naïve Bayes, and Support Vector Machine to detect factors that lead to customer churn. The result showed that promotion seeking was the most factor that caused customer to churn during January and December 2019 period followed by poor customer service, poor network, high cost of services, billing errors, and promotion seeking.

Dassanayake & Herath, 2020 conducted a similar study involving 184 participants obtained through survey method using convenience sampling. Network quality, customer experience, and perceived price were found to have most significant impact on the level of CR.

Hwambo *et al*., 2018 revealed the key factors to churn to be dissatisfaction and poor complaints management system using multiple linear regression analysis.

Thapa, 2018 found network disturbance, low service quality (call quality, call drop), network coverage, data speed, better service offers, better complain resolution and better customer care to be the leading factors of CR.

A similar study was done by (Chen-Hung *et al.*, 2017). The study discovered product involvement, perceived value, transaction convenience, and access convenience being direct predictors of both dependent variables.

Banda & Tembo, 2017 performed a study involving twenty (20) factors where poor customer service (23%), poor network quality (18%), and high cost of services (14%) were the leading factors contributing to churn.

From these studies among others, there seems to be no inconclusive as to which features can be considered as the most influencing towards CR. While some studies did not take demographic details into consideration (Banda & Tembo, 2017; Dassanayake & Herath, 2020; Hwambo *et al.*, 2018; Thapa, 2018), some of the features involving transaction behaviour of customers towards operators have not been exhausted. Furthermore, the mentioned studies among others have been conducted in areas outside Tanzania where operator settings and adopted variables are not exactly similar from the ones adopted in Tanzania. This study intended to determine the features that affect a customer's decision to reduce or increase transactions in mobile operators.

3.0 MATERIALS AND METHODS

3.1 Proposed method

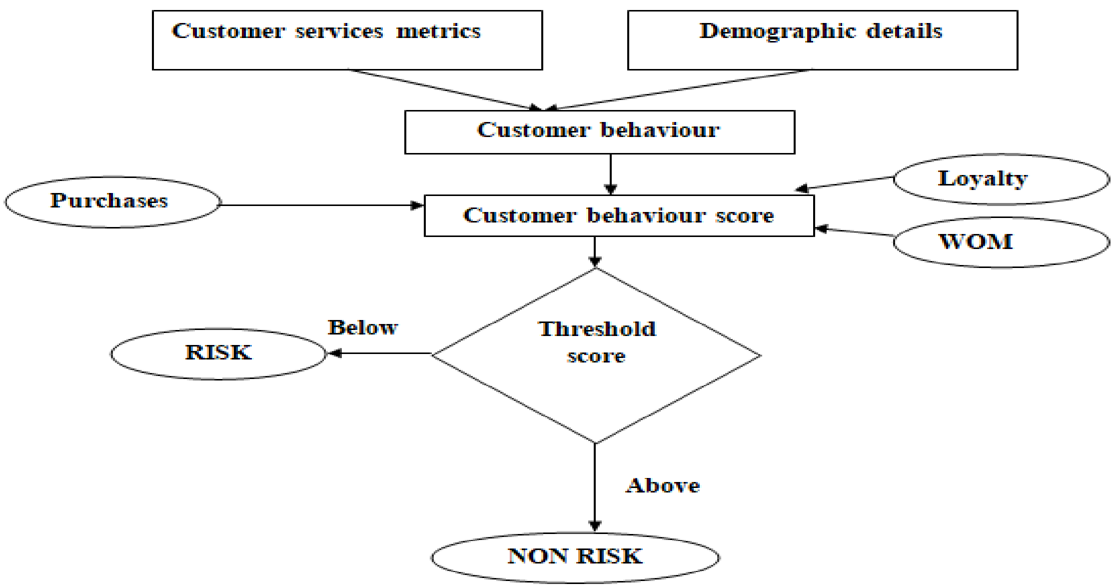

Figure 1: Proposed model of the study

The model included input attributes related to services a customer received from the operator and customers' demographic details. Customer behaviour was considered to be a processed result of the quality of services a customer receives as well as customers' demographic details. The behaviours include how a customer responds to future purchases, customers' loyalty and how customers share the challenges with their colleagues. A combination of various metrics were measured, scored and added to determine the overall behaviour of a customer. The output (RISK or NON_RISK) depended on a set threshold. Threshold value above a set threshold was deemed to be NON RISK customer and below was deemed a customer to be at RISK. This study used a threshold of 3.0 where a score of 3.0 or below was deemed to be RISK since it provided the best balance between RISK and NON_RISK customers. However, a company may tune to any threshold of its choice ranging 0.0 to 6.0. Various machine learning classifications were built to determine which provided the best classification. In this study, risk customers are considered to be the equivalent of potential churners since in non contract based settings which mobile operators adapt in Tanzania, it is difficult to predict churners since majority of them are not removed directly from the database.

3.2 Sampling methods of participants

Purposive sampling method was used to collect data from students. Purposive sampling was used so as to collect data from genuinely interested respondents and avoid non-serious respondents. Respondents were assured that confidentiality of the data will be protected.

3.3 Data collection methods

Questionnaires were administrated to eight (8) university students with a request to fill the questionnaire voluntarily. The questionnaires were distributed to target students who voluntarily were ready to participate. First an introduction was made on what the study was about, and then each question in the questionnaire was elaborated before being filled in to ensure that students understand what they are supposed to answer based on their opinions.

3.4 Ethical considerations

The study maintained ethical standards throughout the study process, obtaining informed consent from university management and survey participants as well as adhering to data protection and privacy regulations.

3.5 Dataset

The initial dataset contained three sets of attributes namely demographic, behaviour and service attributes. The retention attributes were derived from some of the service attributes. The initial lists of attributes are shown in appendix B.

3.5.1Demographic attributes

A total of nine (9) basic demographic attributes were used. However, unique identification attribute, ID was removed as it does not contribute to retention while other attributes DEPENDENTS, AGE, EARNINGS, DEGREE were also eliminated as their categorical members were under represented thus leaving three (3) demographic attributes GENDER, TENURE and LOAN_BOARD to contend with.

3.5.2 Derived demographic attributes

Form the attributes above, LOAN_BOARD beneficiaries were divided into three categories LOW (0.0-0.3), MEDIUM (0.31-0.6) and HIGH (0.61-1.0). Thus a new attribute name called LOAN_BOARD_CLASS was named. Likewise, from REGISTRATION the TENURE attribute was created that shows how long in terms of years a customer has been with the operator.

3.5.3 Behaviour attributes

Behaviour attributes were detected through asking respondents and providing their opinion. Behaviour responses vary based on each customer's interaction and experience with the mobile operator since joining the company to get services.

3.5.4 Customer service attributes

These were students' opinions related to services offered by the operators. Several services are listed under appendix B.

3.5.5 Customer behaviour score

Customers' behaviour score was determined by several metrics as outlined below.

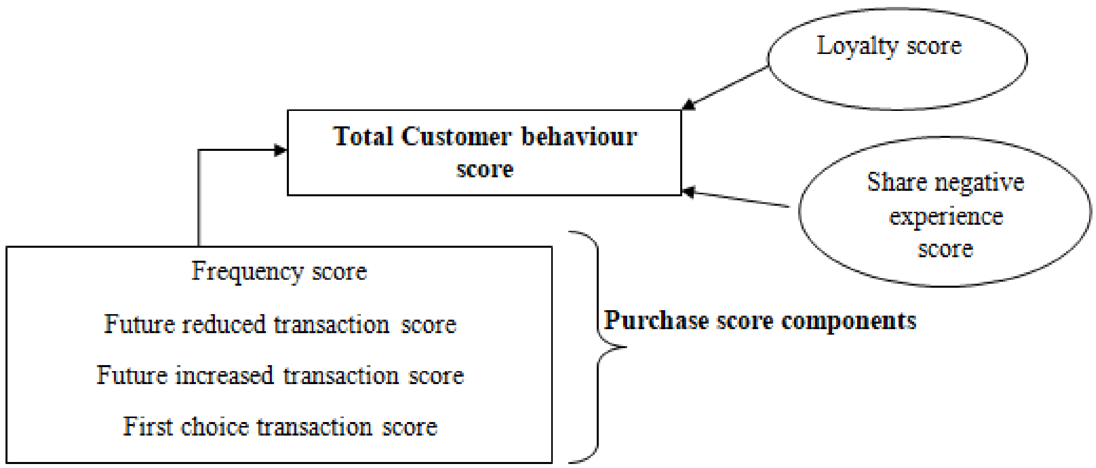

Figure 2: Customer behaviour score metrics

Below is a brief explanation for each of the customer behaviour score components.

Frequency score: This component compares scores obtained from a customer's past transactions score and a customer's expected future transactions. If expected transactions score is greater than past transactions score then Frequency score = 1 meaning that a customer is likely to spend more in the future; if past transactions score is greater than expected transactions score then Frequency scosre = 0 meaning that a customer is likely to spend less in the future and if past transactions score is equal to expected transactions score then Frequency score = 0.5 meaning that a customer is likely to spend equally in the future..

First choice transaction score: This component grades a customer based on a customer's response from the opinion labelled *"This operator is my first choice when it comes to buying bundles"* containing choices Strongly Agree (SA), Agree (A), Fair (F), Disagree (D) and Strongly Disagree (SD). If a customer responds with an 'SA' or 'A' then First choice transaction score = 1 meaning that a customer is likely to spend more in the future; if a customer responds with an 'SD' or 'D' then First choice transaction score = 0 meaning that a customer is likely to spend less in the future and if a customer responds with an 'F' then First choice transaction score = 0.5 meaning that a customer is likely to spend equally in the future.

Future increased transaction score: This component grades a customer based on a customer's response from the opinion labelled *"In the future, I'll buy more bundles from this operator"* containing SA, A, F, D, SD. If a customer responds with an 'SA' or 'A' then Future increased transaction score = 1 meaning that a customer is likely to spend more in the future; if a customer responds with an 'SD' or 'D' then

Future increased transaction score = 0 meaning that a customer is likely to spend less in the future and if a customer responds with an 'F' then Future increased transaction score = 0.5 meaning that a customer is likely to spend equally in the future.

Future reduced transaction score: This component was introduced to act as an offset to the Future increased transaction score for those respondents whose answers are inconsistent. The component grades a customer based on a customer's response from the opinion labelled *"I am likely to reduce transactions from my current service provider to another service provider"* containing SA, A, F, D, SD. If a customer responds with an 'SA' or 'A' then Future reduced transaction score = 0 meaning that a customer is likely to spend less in the future; if a customer responds with an 'SD' or 'D' then Future reduced transaction score = 1 meaning that a customer is likely to spend more in the future and if a customer responds with an 'F' then Future reduced transaction score = 0.5 meaning that a customer is likely to spend equally in the future.

Loyalty score: This component grades a customer based on a customer's response from the opinion labelled *"I am a loyal customer to this operator"* containing SA, A, F, D, SD. If a customer responds with an 'SA' or 'A' then Loyalty score = 1 meaning that a customer is likely to spend more in the future due to customer's; if a customer responds with an 'SD' or 'D' then Loyalty score = 0 meaning that a customer is likely to spend less in the future; and if a customer responds with an 'F' loyalty score = 0.5 meaning that a customer is likely to spend equally in the future.

Share negative experience score: This component grades a customer based on a customer's response from the opinion labelled *"I would also share my negative experience with other customers"* containing SA, A, F, D, SD. If a customer responds with an 'SA' or 'A' then Share negative experience score = 0 meaning that a customer has a negative WOM thus damaging an operator's reputation; if a customer responds with an 'SD' or 'D' then Share negative experience score = 1 meaning that a customer has a positive WOM thus promoting the operator and if a customer responds with an 'F' then Share negative experience score = 0.5 meaning that a customer shares equally in promoting and damaging a operator's reputation.

3.6 Study population

Study population for this study were University students.

3.7 Sample of the study

The sample size was formulated using the following formula:
Sample size $=z^2 *p\times (1-p)/d^2$; where; z^2 = 95% of confidence level and equals 1.96; p = expected prevalence which equals 50%; d^2= level of precision or sampling error and equals 5% (0.05).

Thus, sample size = $1.962\times0.5\times (1-0.5)0.05^2= 384.16 \approx385$.

This sample was adequate for the study (Dash & Malhotra, 2016; Glenn, 1992). However, the larger a sample size is, the more reliable a study is and the better performance a model produces. Hence this study collected 500 questionnaires for getting more reliable results.

4.0 IMPLEMENTATION RESULTS AND DISCUSSION

To determine the features that increase customer retention in mobile operators, python programming was used (Chun, 2001). The initial loaded dataset contained 500 records and 32 attributes. At this point the CLASS output was absent since it was derived later from the TOTAL_SCORE attribute.

After importing the necessary libraries and loading the dataset, data cleaning was performed. At this stage, unnecessary attributes (S/N, MTANDAO and UNIVERSITY) were removed and to avoid losing valuable information, missing values were replaced using modal values for categorical data and median values for numerical data. Thereafter, new attributes were derived from existing attributes and basic attributes that were used to derive other attributes were then dropped.

Output class containing "RISK" and "NON RISK" labels were then determined using the TOTAL SCORE attribute. User was to select the threshold to be used as a cutoff point between a ***'RISK'*** and ***'NON_RISK'*** CLASS student. This study used a threshold of 3.5 where a customer is classified as ***'RISK'*** if a customer's score is 3.5 or less since it gave the best balance between ***RISK*** and ***NON_RISK*** students.

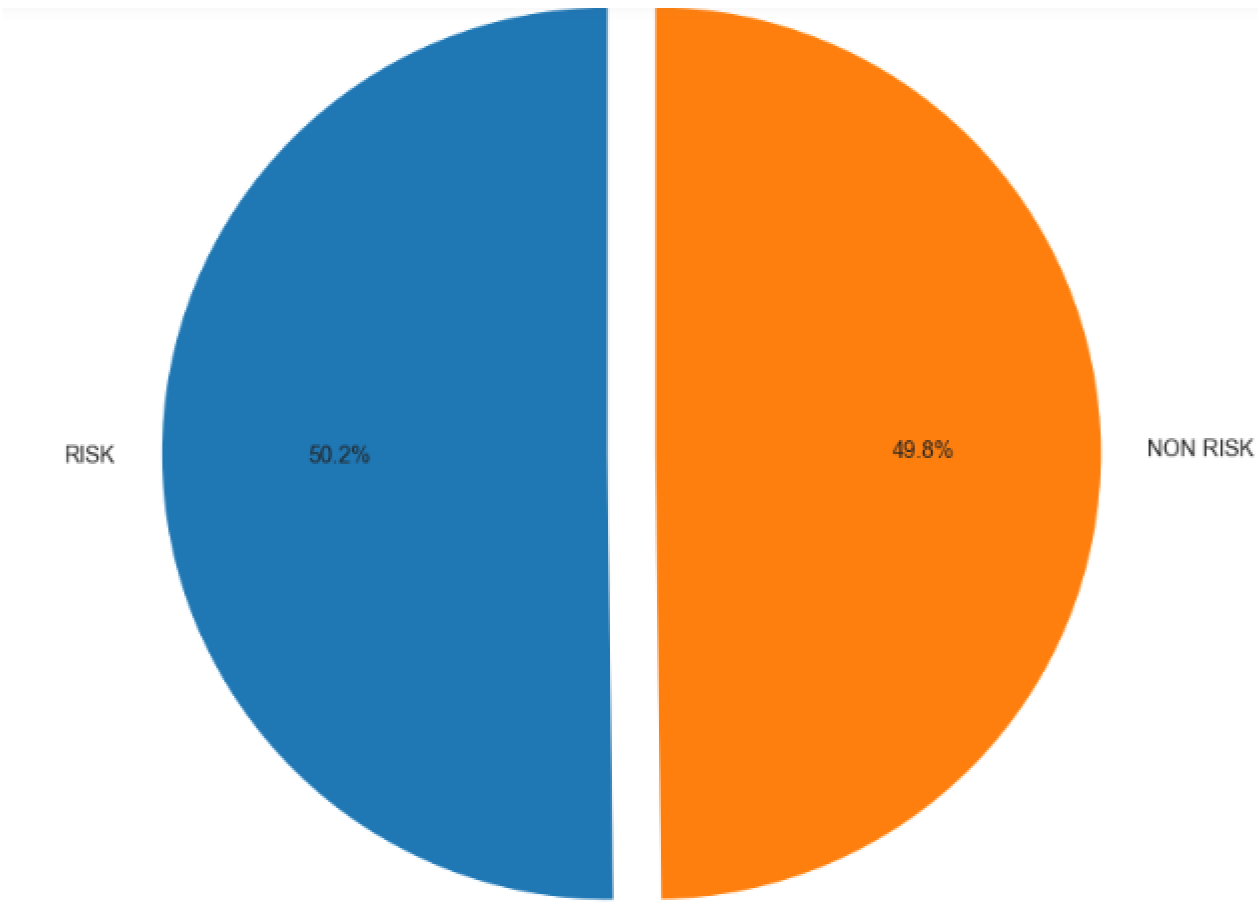

Figure 3: Ratio of RISK and NON_RISK S students at threshold = 3.5

Appendix A indicates ratio of RISK and NON_RISK students for each the remaining thresholds.

Bar plots were created for to determine the ratio of each category in a particular attribute. Personal attributes that are imbalanced were removed as unbalanced attributes reduce the performance capability of machine learning models. Table below shows the distribution of each category from corresponding attribute. From the table above, it can be observed that categories from AGE, MARITAL_STATUS, EARNINGS_CLASS, DEPENDENTS, and DEGREE are unbalanced hence these attributes are removed. The remaining twenty three (23) attributes along with the output class were encoded then score attributes were removed as they were simply used to predict the output class. Feature evaluation from the remaining seventeen (17) features was performed. The impact of each attribute to the output class is shown.

Also the collinearity between features was determined using Variance Inflation Factor (VIF) for seventeen attributes revealed that there was no high VIF among the attributes to affect the models. VIF results are indicated.

```
Variance inflation factor
                      Featureslist       VIF
0                           GENDER  1.937933
1        HIGH_SWITCHING_ENERGY_TIME  2.432551
2               HIGH_SWITCHING_COST  1.981613
3                   VALUE_FOR_MONEY  1.624548
4                  PRICE_COMPARISON  1.937780
5                     WIDE_COVERAGE  2.244795
6                        CALL_DROPS  1.450211
7                    STRONG_SIGNALS  1.658363
8                   PROBLEM_SOLVING  2.478128
9    EXCEPTIONAL_SERVICE_EXPERIENCE  1.835694
10                      GET_THROUGH  2.208754
11                  DO_WHAT_THEY_SAY  3.166568
12       TIMELY_EFFECTIVE_COMPLAINTS  2.176800
13                   FREE_COMPLAINTS  3.226325
14                      TENURE_CLASS  1.511636
15                    EARNINGS_CLASS  1.066102
16                   LOAN_BOARD_CLASS  4.432192
```

Figure 4: Collinearity among features

TABLE 1: Categorical distribution from personal attributes

ATTRIBUTE	**CATEGORY**	**COUNTS**	**PERCENTAGE**	**VERDICT**
GENDER	FEMALE	248	49.6	BALANCED
	MALE	252	50.4	
MARITAL_STATUS	MARRIED	43	8.6	IMBALANCED
	SEPARATED	01	0.2	
	SINGLE	456	91.2	
DEPENDENTS	0 T0 2	400	80.0	IMBALANCED
	3 TO 4	43	8.6	
	MORE THAN 4	57	11.4	
AGE	18-35	492	98.4	IMBALANCED
	36-51	7	1.4	
	51-60	1	0.2	

DEGREE	BACHELOR	464	92.8	IMBALANCED
	CERTIFICATE	9	1.8	
	DIPLOMA	25	5.0	
	MASTERS/PHD	2	0.4	
EARNINGS_CLASS	HIGH	10	2.0	IMBALANCED
	LOW	443	88.6	
	MEDIUM	47	9.4	
LOAN_BOARD_CLASS	HIGH	217	43.4	BALANCED
	LOW	198	39.6	
	MEDIUM	85	17.0	

4.1 Classification model performance

Four models, Decision Tree, Logistic Regression, Random Forest, and AdaBoost were built and tested.and their results are shown as follows.

TABLE 2: Confusion matrix and performance of the classifiers

Models	Confusion Matrix				Performance of classifiers					
	TP	FP	FN	TN	Accuracy (%)		Precision (%)	Recall (%)	F-1 (%)	AUC
					NCV	CV				
Decision Tree	24	23	25	28	52.0	55.0	51.0	49.0	50.0	0.52
Logistic Regression	32	20	17	31	63.0	**64.0**	62.0	65.0	63.0	**0.67**
Random Forest	22	21	27	30	52.0	60.0	51.0	45.0	48.0	0.55
AdaBoost	33	19	16	32	**65.0**	63.0	63.0	67.0	**65.0**	**0.67**

Table shows the AdaBoost model to have the best overall performance where accuracy without Cross validation technique (NCV) and F1-score of 65.0% and 67.0% respectively were achieved. However, Logistic Regression produced the best results in terms of accuracy with CV techniques of 64.0% and shared best result with AdaBoost in terms of Area under Curve (AUC) of 0.67.

Precision refers to the proportion of predicted positive instances that are truly positive (Fujo *et al*., 2022; Hussain *et al*., 2018; Lantz, 2019). In this study, precision was referred to as the proportion of predicted students who will reduce their transactions with an operator and it turns out that these students really were intent on reducing their transactions with an operator.

$$\text{Precision} = [TP] / [TP + FP] \quad (1)$$

It's observed that Adaboost achieved the highest precision indicating that out of 100 students, 63% will be likely correctly attempted to be retained while 37% of those will be likely attempted to be retained while they intend to improve transactions with the operator in the future, thus likely to lose retention bonuses by 37% of the attempted cases.

Recall refers to the number of TPs over the total number of positives (Fujo *et al.*, 2022; Hussain *et al.*, 2018; Lantz, 2019; Sadeghi *et al.*, 2023).

Recall = [TP] / [TP + FN] (2)

It's observed that Adaboost achieved the highest recall indicating that out of 100 students, 67% will be likely correctly attempted to be retained while 33% of those will not be likely attempted to be retained while they intend to reduce transactions with the operator in the future, thus reducing future incomes in the future by 33% of the attempted cases.

4.2 Feature evaluation

Adaboost work well thus it is the best to identify the feature which will increase retention. Here we look at coefficient of each attribute, the greater the coefficient the high the importance.

```
<BarContainer object of 17 artists>
```

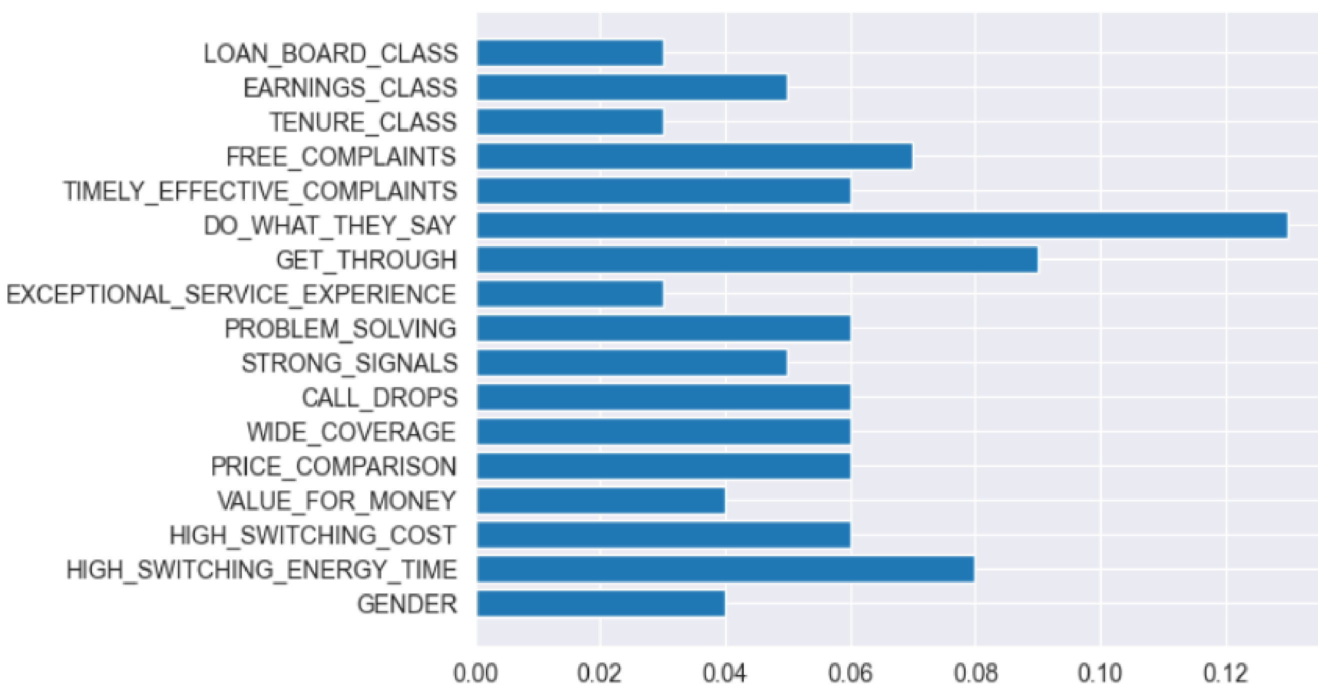

Figure 5: Coefficient of attributes using Adaboost

GSJ: Volume 11, Issue 12, December 2023
ISSN 2320-9186

```
Ranked features:
1. DO_WHAT_THEY_SAY: 0.13
2. GET_THROUGH: 0.09
3. HIGH_SWITCHING_ENERGY_TIME: 0.08
4. FREE_COMPLAINTS: 0.07
5. HIGH_SWITCHING_COST: 0.06
6. PRICE_COMPARISON: 0.06
7. WIDE_COVERAGE: 0.06
8. CALL_DROPS: 0.06
9. PROBLEM_SOLVING: 0.06
10. TIMELY_EFFECTIVE_COMPLAINTS: 0.06
11. STRONG_SIGNALS: 0.05
12. EARNINGS_CLASS: 0.05
13. GENDER: 0.04
14. VALUE_FOR_MONEY: 0.04
15. EXCEPTIONAL_SERVICE_EXPERIENCE: 0.03
16. TENURE_CLASS: 0.03
17. LOAN_BOARD_CLASS: 0.03
```

Figure 6: Ranked features

The study identified features with the most impact on customer transactions in the future were DO_WHAT_THEY, GET_THROUGH, HIGH_SWITCHING_ENERGY_TIME and FREE_COMPLAINTS.

DO_WHAT_THEY_SAY: Students who agreed that their complaints are fulfilled as promised by their operators are more at risk of spending less in the future.

GET_THROUGH: Students who agreed that they easily get through when making calls for complaints are more at risk of spending less in the future.

HIGH_SWITCHING_ENERGY_TIME: Students who agreed that they spent a lot of energy, time and money when they needed an alternative service provider are more at risk of spending less in the future.

FREE_COMPLAINTS: Students who agreed that they receive free complaints in terms of money, energy and time are more at risk of spending less in the future.

5.0 CONCLUSION

In conclusion, this study delved into the intricacies of customer retention within a singular mobile operator, emphasizing the need for cautious generalization to other operators. While the findings offer valuable insights into the identified attributes influencing retention, it is imperative to approach broader industry implications with prudence.

5.1 Recommendations for future studies

Privacy concerns and data collection challenges: Privacy concerns posed a challenge in obtaining detailed customer usage data, prompting the study to rely on self-reported information. To enhance future research, cooperation from operators to share attributes like "PAST_BUNDLES," "PAST_FREQUENCY," and "REGISTRATION" would significantly improve the depth and accuracy of findings. Stringent privacy measures, including supervision and coded data, should be enforced to address concerns.

1. **Expand data Access:**
    - Recommendation*:* Operators are encouraged to collaboratively provide more comprehensive data, including past usage details, for a deeper understanding of customer behavior.
    - Rationale*:* Enhanced access to data will refine analyses and contribute to a more nuanced comprehension of customer retention factors.
2. **Increase sample size:**
    - Recommendation*:* Future studies should consider expanding the sample size to garner additional insights into the effectiveness of machine learning models.
    - Rationale*:* Larger samples enhance the reliability and generalizability of findings, offering a more robust foundation for actionable insights.
3. **Proactive customer retention strategies:**
    - Recommendation*:* Operators are advised to proactively adjust strategies and engage customers falling below retention thresholds before they decide to reduce transaction activities.
    - Rationale*:* Early detection of significant features influencing customer retention allows for timely interventions, fostering long-term customer loyalty.

Involving eight universities and 500 students, this study provides a foundational exploration into customer retention. As we look towards the future, aligning research efforts with these recommendations will contribute to a more comprehensive understanding of customer dynamics, empowering operators to implement targeted strategies and fortify their positions in a competitive market.

6.0 AKNOWLEDGEMENTS

This study acknowledges the management of universities that granted access to their students for data collection, along with the University students for their willingness to participate in the study and provide crucial details that aided in data analysis and building the necessary models.

7.0 DECLARATIONS

7.1 Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

7.2 Ethics approval and consent to participate

Stated in section 3.4

7.3 Consent for publication

Not applicable

7.4 Availability of data and materials

The dataset analyzed during the current study is available from the corresponding author on reasonable request.

7.5 Competing interests

The authors declare that they have no competing interests.

7.6 Funding

Funds used for this research came from Sokoine University of Agriculture as part of sponsorship for my PhD studies.

REFERENCES

Ahn, J.-H., Han, S.-P., & Lee, Y.-S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, *30*(10–11), 552–568. https://doi.org/10.1016/j.telpol.2006.09.006

Alshurideh, M. (2016). *Scope of Customer Retention Problems in the Mobile Phone Sector: A Theoretical Perspective*.

Amatare, S. A., & Ojo, A. K. (n.d.). *Predicting Customer Churn in Telecommunication Industry Using Convolutional Neural Network Model*.

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, *94*, 290–301.

Bahri-Ammari, N., & Bilgihan, A. (2017). The effects of distributive, procedural, and interactional justice on customer retention: An empirical investigation in the mobile telecom industry in Tunisia. *Journal of Retailing and Consumer Services*, *37*, 89–100. https://doi.org/10.1016/j.jretconser.2017.02.012

Bahri-Ammari, N., & Bilgihan, A. (2019). Customer retention to mobile telecommunication service providers: The roles of perceived justice and customer loyalty program. *International Journal of Mobile Communications*, *17*(1), 82–107.

Banda, P., & Tembo, S. (2017). Factors Leading to Mobile Telecommunications Customer Churn in Zambia. *International Journal of Engineering Research in Africa*, *31*, 143–154. https://doi.org/10.4028/www.scientific.net/JERA.31.143

Cheng, B. L., Gaur, S. S., & Rezuan, A. R. (2020). Factors leading to customer retention in the high volume-low volume service context: Evidence from the mobile service industry. *Asian Journal of Business Research*, *10*(1).

Chen-Hung, T., Chin-Chiung, K. U. O., & Marianne, J. E. (2017). The antecedents and consequences of word of mouth: A meta-analysis. *Journal of Social and Administrative Sciences*, *4*(2), 181–192.

Chopra, C., Gupta, S., & Manek, R. (2020). Impact of social media on consumer behaviour. *International Journal of Creative Research Thoughts*, *8*(6), 1943–1961.

Chun, W. (2001). *Core python programming*. Prentice Hall.

Dash, M., & Malhotra, N. K. (2016). Marketing Research an Applied Orientation. *Indian Institute of Management Lucknow*.

Dassanayake, D., & Herath, R. (2020). *Factors Affecting on Customer Retention in Mobile Telecommunication Industry in Sri Lanka*.

Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, *11*(1), 24.

Glenn, D. I. (1992). Determining sample size. *A Series of the Program Evaluation and Organizational Development. University of Florida, Publication Date: November*.

Hanaysha, J. (2016). Restaurant Location and Price Fairness as Key Determinants of Brand Equity: A Study on Fast Food Restaurant Industry. *Business and Economic Research*, *6*, 310. https://doi.org/10.5296/ber.v6i1.9352

Hassan, S. (2013). *Measuring customer satisfaction and loyalty through service fairness, service quality and price fairness perception: An empirical study of Pakistan mobile telecommunication sector*.

Hussain, S., Abdulaziz Dahan, N., Ba-Alwi, F. M., & Ribata, N. (2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447. https://doi.org/10.11591/ijeecs.v9.i2.pp447-459

Hwambo, L., Shamhuyenhanzva, R. M., & Sandada, M. (2018). An Assessment of Churn Determinants in Zimbabwe'S Mobile Telecommunication Services. *International Journal of EBusiness and EGovernment Studies*, *9*(2), 106–120.

Izogo, E. E. (2017). Customer loyalty in telecom service sector: The role of service quality and customer commitment. *The TQM Journal*, *29*(1), 19–36.

Kaur, S. (2017). Literature Review of data mining techniques in customer churn prediction for telecommunications industry. *Journal of Applied Technology and Innovation*, *1*(2), 28–40.

Lancaster, H. (2020). *Tanzania—Telecoms, Mobile and Broadband—Statistics and Analyses—BuddeComm*. Retrieved August 23, 2023, from https://www.budde.com.au/Research/Tanzania-Telecoms-Mobile-and-Broadband-Statistics-and-Analyses

Lantz, B. (2013). *Machine Learning with R*. Packt. https://edu.kpfu.ru/pluginfile.php/278552/mod_resource/content/1/MachineLearningR__Brett_Lantz.pdf

Lemon, K. N., White, T. B., & Winer, R. S. (2002). Dynamic customer relationship management: Incorporating future considerations into the service retention decision. *Journal of Marketing*, *66*(1), 1–14.

Nasir, A., Mushtaq, H., & Rizwan, M. (2014). Customer loyalty in telecom sector of Pakistan. *Journal of Sociological Research*, *5*(1), 449–467.

Oblander, E. S., Gupta, S., Mela, C. F., Winer, R. S., & Lehmann, D. R. (2020). The past, present, and future of customer management. *Marketing Letters*, *31*, 125–136.

O'Dea, S. (2020). *US wireless telecom carriers churn rate 2013-2018 | Statista*. https://www.statista.com/statistics/283511/average-monthly-churn-rate-top-wireless-carriers-us/

Rahman, M. B., & Chowdhury, N. (2022). Determinants of customer retention in cellular industry of Bangladesh: A study on young customers. *Indonesian Journal of Social Research (IJSR)*, *4*(1), 10–21.

Rather, R. A. (2019). Consequences of consumer engagement in service marketing: An empirical exploration. *Journal of Global Marketing*, *32*(2), 116–135.

Sadeghi, M., Dehkordi, M. N., Barekatain, B., & Khani, N. (2023). Improve customer churn prediction through the proposed PCA-PSO-K means algorithm in the communication industry. *The Journal of Supercomputing*, *79*(6), 6871–6888.

Saputro, B., Ma'mun, S., Budi, I., Santoso, A. B., & Putra, P. K. (2021). Customer churn factors detection in Indonesian postpaid telecommunication services as a supporting medium for preventing waste of IT components. *IOP Conference Series: Earth and Environmental Science*, *700*(1), 012015.

Schmittlein, D. C., & Peterson, R. A. (1994). Customer base analysis: An industrial purchase process application. *Marketing Science*, *13*(1), 41–67.

Shaaban, E., Helmy, Y., Khedr, A., & Nasr, M. (2012). A Proposed Churn Prediction Model. *International Journal of Engineering*, *2*(4).

Tan, Q., Oriade, A., & Fallon, P. (2014). *Service Quality and Customer Satisfaction in Chinese Fast Food Sector: A Proposal for CFFRSERV Scale*.

TCRA. (2019). *TCRA. (2019). Quarterly communications statistics. Https://www.tcra.go.tz/uploads/text-editor/files/september_1619156609.pdf—Google Search*. https://www.google.com/search?hl=en&q=TCRA.+(2019).+Quarterly+communications+statistics.+https://www.tcra.go.tz/uploads/text-editor/files/september_1619156609.pdf

TCRA. (2020). Quarterly communications statistics. Https://www.tcra.go.tz/uploads/text-editor/files/december%20(1)_1619157631.pdf. (n.d.).

Thapa, P. (2018). An Analysis of the Implementation of the Factors of Customer Retention by Nepal Telecom. *International Journal of Social Sciences and Management*, *5*(3), 89–97.

Townsend, A., & Nilakanta, S. (2019). *Customer churn: A study of factors affecting customer churn using machine learning* [PhD Thesis]. Iowa State University.

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, *7*, 60134–60149.

Umayaparvathi, V., & Iyakutti, K. (2012). Applications of data mining techniques in telecom churn prediction. *International Journal of Computer Applications*, *42*(20), 5–9.

Verhoef, P. C., Franses, P. H., & Hoekstra, J. C. (2002). The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider: Does age of relationship matter? *Journal of the Academy of Marketing Science*, *30*, 202–216.

APPENDIX

Appendix A: Relationship between RISK and NON_RISK students at given thresholds

S/N	THRESHOLD	NON_RISK (%)	RISK (%)
1.	0.0	98.6	1.4
2.	0.5	98.0	2.0
3.	1.0	95.4	4.6
4.	1.5	92.6	7.4
5.	2.0	86.0	14.0
6.	2.5	77.2	22.8
7.	3.0	62.2	37.8
8.	**3.5**	**49.8**	**50.2**
9.	4.0	30.8	69.2
10.	4.5	20.0	80.0
11.	5.0	5.0	95.0
12.	5.5	2.0	98.0
13.	6.0	0.0	100.0

Appendix B: List of selected attributes

S/N	VARIABLE NAME	DESCRIPTION
DEMOGRAPHIC DETAILS	GENDER (Cheng *et al*., 2020)	Customer's gender, {Male, Female}
	MARITAL_STATUS (Cheng *et al*., 2020)	{married, single}
	DEPENDENTS	Number of people under the custody of a customer
	AGE (Cheng *et al*., 2020)	Age range of a customer, {18-35, 36-50, 51-60, above 60}
	EARNINGS (Cheng *et al*., 2020)	Customer's monthly earning range.
	EARNINGS_CLASS	Derived attribute from earnings {HIGH, LOW, MEDIUM}
	REGISTRATION (Verhoef *et al*., 2002)	Year customer joined the operator
	LOAN_BOARD	Loan board percentage received by a customer or its equivalent

	LOAN_BOARD_CLASS	Derived attribute from LOAN_BOARD {LOW, MEDIUM, HIGH}
RETENTION DETAILS	PAST_BUNDLES (Lemon *et al*., 2002; Schmittlein & Peterson, 1994)	Estimated bundles utilized by a customer in the last three months.
	PAST_FREQUENCY (Lemon *et al*., 2002; Schmittlein & Peterson, 1994)	Estimated number of bundles utilized by a customer in the last three months.
	FUTURE_BUNDLES (Lemon *et al*., 2002; Schmittlein & Peterson, 1994)	Expected bundles to be utilized by a customer in the next three months.
	FUTURE_FREQUENCY (Lemon *et al*., 2002; Schmittlein & Peterson, 1994)	Expected number of bundles to be utilized by a customer in the next three months.
	SCORE_FREQUENCY	Derived attribute which compares past transaction frequencies against future transaction frequencies, {0,0.5,1}
	SHARE_NEGATIVE_EXPERIENCE (Izogo, 2017)	Customer discusses challenges with others, {Agree, Disagree, Fair}
	SCORE_SHARE_NEGATIVE_EXPERIENCE	Derived attribute from SHARE_NEGATIVE_EXPERIENCE, {Agree =0, Disagree =1, Fair=0.5}
	REDUCE_TRANSACTIONS(Bahri-Ammari & Bilgihan, 2017, 2019; Hassan, 2013)	Customers will reduce doing transactions with the operator, {Agree, Disagree, Fair }
	SCORE_REDUCE_TRANSACTION	Derived attribute from REDUCE_TRANSACTIONS, {Agree =0, Disagree =1, Fair=0.5}
	FUTURE_TRANSACTIONS(Bahri-Ammari & Bilgihan, 2017, 2019; Hassan, 2013)	Customer is willing to do future transactions with the operator. {Agree, Disagree, Fair}
	SCORE_FUTURE_TRANSACTION	Derived attribute from FUTURE_TRANSACTIONS, {Agree =1, Disagree =0, Fair=0.5}
	LOYAL (Bahri-Ammari & Bilgihan, 2017, 2019; Hanaysha, 2016; Hassan, 2013)	Customer is loyal to the operator. {Agree, Disagree, Fair}
	SCORE_LOYAL	Derived attribute from LOYAL, {Agree =1, Disagree =0, Fair=0.5}

	FIRST_CHOICE (Bahri-Ammari & Bilgihan, 2017, 2019)	This is customer's first choice operator when deciding a transaction, {Agree, Disagree, Fair}
	SCORE_FIRST_CHOICE	Derived attribute from FIRST_CHOICE {Agree =1, Disagree =0, Fair=0.5}
	TOTAL_SCORE	Derived attribute that adds all the derived scores, SCORE_frequency + SCORE_future_TX + SCORE_reduce_TX + SCORE_loyal + SCORE_first_choice
	CLASS	Predictor derived from TOTAL_SCORE, {RISK, NON_RISK}
SERVICES	GET_THROUGH	Customer is able to immediately get in touch with care personnel for enquiry, {Agree, Disagree, Fair}
	DO_WHAT_THEY_SAY (Izogo, 2017; Tan *et al.*, 2014)	Service personnel keep their promise on what they say, {Agree, Disagree, Fair}
	TIMELY EFFECTIVE COMPLAINTS (Bahri-Ammari & Bilgihan, 2019)	Student's opinion on how long complaints are solved, {Agree, Disagree, Fair}
	FREE COMPLAINTS	Complaints are free of charge, {Agree, Disagree, Fair}
	HIGH_SWITCHING_ENERGY_TIME (Thapa, 2018)	A lot of energy and time is utilized when applying for a new line, {Agree, Disagree, Fair}
	HIGH_SWITCHNG_COST (Thapa, 2018)	High monetary cost is involved when applying for a new line, {Agree, Disagree, Fair}
	VALUE_FOR_MONEY (Hanaysha, 2016; Hassan, 2013)	Services offered and money paid is value for money, {Agree, Disagree, Fair}
	CHEAP_PRICE_COMPARISON (Hassan, 2013)	Customer's opinion in comparing primary operator services vs secondary operator services. {Agree, Disagree, Fair}
	WIDE_COVERAGE (Nasir *et al.*, 2014)	Customer's opinion on the operator's network coverage wherever a customer goes, {Agree, Disagree,

		Fair}
	STRONG_SIGNALS	Customer's opinion on the network signals, {Agree, Disagree, Fair}
	PROBLEM_SOLVING (Bahri-Ammari & Bilgihan, 2019)	Customer's opinion on how problems are handled, {Agree, Disagree, Fair}
	EXCEPTIONAL_SERVICE_EXPERIENCE (Hassan, 2013)	Customer's opinion on the services received, {Agree, Disagree, Fair}