Context-Aware Computational Trust Model for Recommender Systems

Edwin O. Ngwawe, Elisha O. Abade, and Stephen N. Mburu

Abstract — With increase in computing and networking technologies, many organizations have managed to place their services online with the aim of achieving efficiency in customer service as well as reach more potential customers, also with communicable diseases such as COVID-19 and need for social distancing, many people are encouraged to work from home, including shopping. To meet this objective in areas with poor Internet connectivity, the government of Kenya recently announced partnership with Google Inc for use of Google Loon. This has come up with challenges which include information overload on the side of the end consumer as well as security loopholes such as dishonest vendors preying on unsuspecting consumers. Recommender systems have been used to alleviate these two challenges by helping online users select the best item for their case. However, most recommender systems, especially common filtering recommendation algorithm (CFRA) based systems still rely on presenting output based on selections of nearest neighbors (most similar users – birds of the same feathers flock together). This leaves room for manipulation of the output by mimicking the features of their target and then picking malicious item such that when the recommender system runs, it will output the same malicious item to the target – a trust issue. Data to construct trust is equally a challenge. In this research, we propose to address this issue by creating a trust adjustment factor (TAF) for recommender systems for online services.

Index Terms — online ethics, online shopping, recommender systems, trust modeling for computational purposes.

I. INTRODUCTION

Over the years there have been significant advances in Information and Communication Technologies (ICT) resulting in increased availability of high speed, reliable broadband internet services. Further, there has been consistent reduction in prices of computing gadgets such as smartphones and tablets. These factors have resulted in increased Internet penetration and utilization of online services.

The growing number of younger population of Internet users is also a factor that has accelerated uptake of online services, especially in developing economies such as Kenya. This has seen the disruption of the supply chain and logistics as previously known and a growing acceptance of online shopping, social media services and telemedicine. A recent survey in Kenya showed that people would most likely

search online (on Google) for diagnosis services or information about an item they intend to purchase.

Recently, the COVID-19 pandemic has also come with its own significant disruption of life as we have known it. This has resulted in a significant push of several services online. Due to COVID-19 WHO gave guidelines on maintaining hygiene, keeping social distancing among other things (World Health Organization. Social distancing has resulted in pushing of a significant number of traditional services to the online space. These include learning, shopping, medical consultation, gaming, hotel and restaurant services and even other forms of entertainment such as gaming and gambling.

Nonetheless, this does not eliminate some of the pre-COVID-19 challenges that have afflicted the online services. With the growing speeds and computing power of Information technology [4], [17], the connectivity and integration of various computing services have emerged. The amount of data being processed at any given time is huge in volume, leading to information overload. Therefore, one of the challenges has been the difficulty in selecting a reliable online service due to information overload [10]. To solve this, recommender services have been proposed. A recommender system can be defined as a tool that helps an online user to choose services online against a myriad of options. Recommender systems can help to alleviate the burden of decision making from users [10]. Recommender systems are implemented using numerous technologies such as collaborative filtering, content-based filtering and latent factor filtering.

In as much as collaborative filtering is seen as a successful technique for recommender systems [7], it is still possible for malicious vendors to manipulate the outcome of a recommender system unfairly by providing false historical data to the recommender systems or using other digital marketing techniques to their advantage albeit unfairly hence leaving a room for distrust in these systems [13]. According to [18], there has been a gradual increase in cyber attacks made by attacking vulnerabilities in software systems. This has led to distrust of online and recommender systems in general since they are susceptible to manipulation.

This distrust does not only deny the online users the advantages of online services such as reduced average prices of online shops, time saving as one shops conveniently from home and also the cost of locomotion but also disrupts the potential gains of social distancing during a communicable pandemic hence leading to exposure to potential public health danger.

In this paper, we propose model that can be used to predict trustworthiness of an online services using Structural Equation Modeling (SEM). The proposed model can predict

Published on November 9, 2020.

E. O. Ngwawe, School of Computing and Informatics, University of Nairobi, Kenya.

(corresponding e-mail: Edwin.ngwawe@uonbi.ac.ke)

E. O. Abade, University of Nairobi, Kenya.

(e-mail: eabade@uonbi.ac.ke)

S. N. Mburu, University of Nairobi, Kenya.

(e-mail: snmburu@uonbi.ac.ke).

the trustworthiness of an online service and help adjust the output of automated recommender systems by filtering out suspicious services in a context aware manner.

The rest of this paper is organized as follows. In section two, we discuss some of the previous research works that are closely related to this study. In section three we present our proposed solution then discuss preliminary results in section four. We then conclude and give areas of future works in section five.

II. RELATED WORK

A. Overview of Trust

Trust is the substantiated belief that an agent will comply with the expected standards in a given context and deliver desired results. It is a latent construct that cannot be measured directly but through some indicators which are based on the said context. Many sources in literature associate trust closely with ethics. In online shopping experience, it is the shopper's desire to get the item he is buying for in the promised form, at the right place and at the promised price and promised time. However, some unscrupulous online vendors exploit the obscurity of online systems, in the sense that the buyer does not have the full view of the promised item at the time of purchase and therefore dupe the buyer. Traditionally, there are indicators that buyers use to discern fraud and make informed decision in the normal brick and mortar shop, but at an online shop, the buyer is limited only to the information the seller has provided to him about the item or service, and therefore the buyer is at the mercy of the seller, to some extent.

B. Trust Measurement Methods

1) Scale Development using Confirmatory Factor Analysis

Factor analysis is a regression method that is applied to discover root causes that explain hidden factors that are present in data. It is a method used to explore datasets to find out why data is acting in a certain way. Factors are also known as latent variables or constructs, that is, variables that are quite meaningful but are inferred and not directly observable. For example, imagine you are a marketing data scientist and that you must add to a file actionable customer segments for use in strategic marketing planning. You have got a response from customer survey. You can apply factor analysis to group respondents into meaningful customer segments based on similarities on how responses tend to answer a specific subset of survey questions. So factor analysis is a method that you can use to regress on features in order to discover factors that you can use as variables to represent the original dataset. It is important to note that factor analysis is a two step process which involves:

- (i) Exploratory Factor Analysis (EFA).
- (ii) Confirmatory Factor Analysis (CFA).

The main difference is that for the Exploratory Factor Analysis, we are keen on reducing the variables into few meaningful factors. Here the factor loadings are calculated, which determine what amount of relationship exists between a variable and a certain factor. Here we rotate the factors to get a nice distribution so that the variables are not loaded into one factor but they are distributed across several factors.

Here we can have an example of orthogonal rotation where we assume that variables are uncorrelated; we can also have oblique type of rotation where we assume that the variables have some degree of correlation among them. Here, the main task is to explore relationships between exogenous and endogenous variables (indicators vs. factors or latent variables or constructs). Once we have these factors, say F_1, F_2, F_3, F_4 , we can use the values derived from each of them for some kind of predictive modeling like in a multiple regression.

For the case of Confirmatory Factor analysis, as the name suggests, the researcher already has some prior knowledge about constructs and the variables in terms of which variable will make up to which factor or a construct and has some theoretical foundation about that so here, the task is not to explore any more but to confirm that whatever was thought of is actually true just as quality measure.

The two steps above (EFA and CFA) are also referred to as measurement model in Structural Equation Modeling, discussed later in this paper.

Research work such as [11] has studied the ethics on online retail in European context using the measurement models described above (CFA and EFA) and has provided some factors that shoppers are worried about which implies trust, alongside the variables or indicators that the factors are inferred from. The factors are:

- (i) Security;
- (ii) Reliability;
- (iii) Fear of deception;
- (iv) Privacy.

These factors can be used for scientific prediction of trust modeling. Since the community norms are different from context to context and from continent to continent, it is very likely to realize that the indicators will be very different in another context or continent, since many of the clues that are used to detect fraudulent activities are usually based on previous experiences and insights that have been passed from one generation to the other over hundreds of years and these vary greatly from community to community. The significance of this will be to say appeal to all communities by instilling confidence in them, from their perspectives, by taking care of how they evaluate trust in the online shops or any other context.

2) Structural Equation Modelling (SEM)

Structural Equation Modeling involves creating equations which depict relationships among constructs involved in some analysis. Structural Equation Modeling comprises of three important parts, namely:

- (i) Factor analysis;
- (ii) Regression Analysis;
- (iii) The Chi Square value, which is largely used to test the goodness of fit.

Again, the constructs are unobservable and can only be measured through some items or variables in the questionnaire. A construct can be measured by any number of items in the questionnaire, but a researcher need to be careful not to take very few variables as this will lead into a situation where a model cannot properly explain itself, a shortcoming known as under fitting or low bias, in other words, if to be applied in artificial intelligence, then a machine learning model cannot fit the training data or

generalize to new data. Again using so many variables can lead to over fitting or high variance of the model, where a model can produce almost 'accurate' performance in the training data, but does not generalize, such that it cannot produce accurate results in unforeseen data and this is against our objective (International Business Machines Corporation (IBM), 2019).

So, in structural equation modeling, we create two models, namely:

- (i) Measurement Model;
- (ii) Structural model.

In the measurement model, one measures whether the variables are actually measuring the constructs, or not, using the EFA and CFA discussed in section B (1) above.

In Structural model, we seek to see the structure of the relationships.

Here we have several types of relationships that we need to assess so as to form an equation that can be used for mathematical prediction of latent variables or constructs, something in the form of $y = mx + c$, where y is the desired construct to be predicted, m is the weight and coefficient of the relationship and c is some constant such as error constant.

The types of the relationships are:

- (i) A relationship between a construct and a measured variable, which can be exogenous or endogenous.

- (ii) Relationship between a construct and multiple measured variables.

- (iii) Dependence Relationship Between Two Constructs (Structural Relationship)

- (iv) Correlational Relationship Between Constructs

A measurement model can be represented by types I, II, IV.

While a structural model includes all the types of relationships.

A structural model is a one where a researcher is basically testing a set of constructs with an intention to measure the relationships and how it affects and try to determine the path estimates and from there he can come to some kind of a conclusion or an inference. With this type of modeling, trustworthiness of a service can be inferred or predicted computationally. On the other hand, measurement model we only check if the variables are actually measuring the construct. So in measurement model, we say that the variables are actually measuring the construct and in structural model, we say that there is a relationship and we are trying to find that relationship.

Since SEM explains the observed covariance among a set of measured variables, by estimating the observed covariance matrix with estimated covariance matrix. (The estimated covariance matrix is constructed based on the estimated relationships among the variables), it is desirable that the difference is as small as possible to ensure that what is observed vs. what is desirable are more or less the same. We use chi square to test this.

$$\chi^2 = \frac{(\text{Observed Matrix (O)} - \text{Expected Matrix (E)})^2}{\text{Expected Matrix (E)}}$$

With the above equation, we can see that chi square becomes large in two cases:

- (i) When the difference between observed value (o) and expected value (E) is so large.
- (ii) When the expected value E is just so small.

In these two cases, we will understand that there is a significant difference between observed and the estimated model, and this is against our wish, since we do not want a very high difference between the estimated and the observed value so in SEM, a low chi square value is more desired as it means a better fitting model while a high chi square value implies a poorly fit model.

It is important to note that no SEM model should be developed without any underlying theory, since the SEM software, Mostly AMOS graphic, will still give you some results with whichever data but if the results cannot be founded on any theory, then it is not important. The basis of SEM is always some theory unless there is a scientific reason that the theory still needs to be developed, in which case it is also good thing but be done with care. This second scenario is important as it can contribute to the body of knowledge, otherwise any researcher should be discouraged from using SEM without a proper thought process.

To define individual constructs, the following steps need to be taken:

- (i) Operationalization of the construct;
- (ii) Use of scale from prior research;
- (iii) New scale development;

(iv) Pre-testing of the construct.

To define the individual constructs, as said earlier, one needs to go with some theory. That is to say, they need to understand what the construct is and why it is required and what variables are or would affect the construct. Only when these questions are answered can one then develop a construct. Then one needs to support it with sufficient literature and research.

Once one has defined the construct, then one can use a new scale, but many at times, even in the cases of confirmatory factors, scales from previous researches are used since such a scale usually has been tested and confirmed or validated somewhere else. So if you are a researcher who is trying to use this construct for some other study (but it has to be theoretically sound) then scales from prior research can be taken and again checked for validation.

As said earlier, a new scale can be developed and validated. This method is always appreciated because it results into adding more to the body of knowledge or is like the researcher is contributing to the knowledgebase.

Once the construct has been defined and a scale is developed, then the construct needs to be pre-tested. We need to test whether the variables are loading to the constructs properly or not. For example if you had taken, say, five variables then you test and realize that either one or just some, or even all of them are not loading properly onto the construct, then there is no point of going ahead with that scale. Also the variables could be cross-loading and this is equally dangerous so again in this case you will need to drop the scale.

So under SEM, the researcher works in two stages. First he checks the measurement model – if the variables measure a construct as checked through factor loadings and the extent of measurement errors and then he checks the structural model where he focuses on relationships of the construct as given by the structural model.

3) Other Trust Measurement models

Other sources show that trust measurement can also be modeled using:

- (i) Using open network environment [1].
- (ii) Measuring trust using agent synergy [15].
- (iii) Using Techniques of Social Trust [16].
- (iv) Using Techniques of Game theory [15].
- (v) Using Techniques of Psychology [5].

C. Trust and Recommender Systems

Recommender systems purely learn from provided data to make predictions. Now what will happen when a selfish and malicious member of the online community deliberately provides falsified data? Clearly, left unchecked, the output of recommendation will subsequently be falsified and depending on the area of application, the results can be catastrophic. Take an example of e-tourism application where a malicious user can perpetrate social engineering by providing malicious input to the recommender system which may lead the recommender system to recommend a certain destination to a prospecting tourist. Left unchecked, this location can be a dangerous location that if the tourist visits, then he may fall a prey to the perpetrator's interest which may mean loss of life, property or other form of physical or psychological harm. In short, output of a recommender

system can be manipulated and as such they are not trustworthy in their natural state.

D. Possibility of Incorporating Trust into Recommender Systems

Even though trust seems an amorphous construct that cannot be physically measured, a scientific paper [18], has shown that if it is incorporated into a recommender system then it improves the most desirable property of recommender system – the accuracy as measured by mean absolute error (MAE) and Root Means Square Error (RMSE). This is as demonstrated by the graphs below which depict their results. The graphs indicate the relationship between precision of a recommender algorithm measured by Mean Absolute Error (MAE) for Figure 1 and by Root Mean Square Error (RMSE) for Fig. 2.

As can be seen, the accuracy is compared against the same number of neighbors for all cases. In Common Filtering Recommendation Algorithm, neighborhood means the most similar users or items to the item or user in question, in a given set of items or users under review.

It is a paramount to intuit that the more the neighbors, other factors remaining the same, the better the prediction since more neighbors imply more data to learn a better pattern for prediction from. In this case, this was held constant.

Fig. 1. Accuracy results comparison of variants of recommender systems using MAE (source: [18]).

As a matter of fact, choosing the 'proper' number of neighbors is a topic of scientific research on its own [10].

Here the study looks at the Collaborative Filtering Recommendation Algorithm (CFRA), Common Filtering Recommendation Algorithm with Trust incorporated, (CFRAT) and Hybrid Recommendation Algorithm, with Trust (HRAT).

Collaborative Filtering Recommendation Algorithm (CFRA) is a user – user filtering, it works in the following manner: considering a user x , we to find a set N of users whose ratings are similar to user x ratings. We then estimate user x 's ratings based on the ratings of users in set N . We call this set of users in N the Neighbourhood of user N .

Common Filtering Recommendation Algorithm with Trust incorporated, (CFRAT) is a new concept of incorporating trust to a recommender system.

Fig. 2. Accuracy results comparison of variants of recommender systems using RMSE (source: [18]).

Hybrid Recommendation Algorithm, with Trust (HRAT) is a recommendation algorithm which uses both Content-based approach, discussed in section and Collaborative Filtering Recommendation algorithm discussed combined.

As can be seen from the graphs, it is evident that:

(i) the more the neighbors, the more the error goes down, in other words the accuracy improves.

(ii) When trust is incorporated, again the error goes down, measured by both mean absolute error (MAE) and root means Square Error (RMSE).

(iii) Hybrid Recommendation algorithm incorporated with trust (Trust adjustment factor) or trust enhanced Hybrid recommended algorithm (HRAT) outperforms trust enhanced common filtering recommendation algorithm (CFRAT) which in turn outperforms the plain collaborative filtering recommendation algorithm (CFRA) which is in its natural state

In other words, trust should be a key element of recommender systems.

However, this study paper [18] was carried out from a dataset where users explicitly indicated their trust levels against other users in the system in a publicly available opinions website.

This does not only require extra efforts from the users as discussed in section 3 and also is susceptible to cold start problems, a problem where the recommender system is unable to provide accurate recommendations because it has not yet gotten enough data to learn patterns from.

It is also worth noting that the data set that was used in this study [18] is no longer available to researchers as the website is no longer there and the data that was downloaded earlier from the website can now be considered old.

The discussion in this section necessitates the need to come up with new mechanisms of acquiring the trust constructs for the purpose of computation for recommender system algorithms.

III. PROPOSED SOLUTION

As can be seen from the ongoing discussions, trust adjustment factor improves the precision of a recommender system, and also can help filter out fraudulent services. But how do get to quantify the trust element itself scientifically for the purpose of scientific computation?

We are using to use Structural Equation Modeling (SEM) to estimate the trustworthiness of an online service in a context aware manner.

We have carried carry out a research in Kenya, a typical context of African country, and found out the measurable variables that construe trust from the perspective Kenyan online users.

The research involves the following stages:

- (i) Item generation;
- (ii) Exploratory Factor Analysis;
- (iii) Confirmatory Factor Analysis;
- (iv) Structural Equation Modeling;
- (v) Experimental tests.

A. Item Generation

To get the items for the study, we first adopted the items from previous studies [11]. We then reviewed the items in 11 in-depth interviews and 5 focused group discussions, and the purpose of these reviews was to:

(i) Help in the process of defining the dimensions of the construct.

(ii) Generate new items.

(iii) Perform a thorough evaluation of the item wording.

(iv) Eliminate any redundant, ambiguous, or poorly worded items.

The in-depth interviews compromised members of Faculty in leading universities in Kenya.

The number of in-depth interviews and focused group discussions were arrived at after attainment of theory saturation, a state where we were not learning any more information from the respondents.

The focused group discussions comprised of seven members, each. The members were conveniently sampled from members of a university community, comprising of faculty and non-faculty members and students, some who had never purchased any item online while other had purchased varied number of items online over a period of time. The composition of individual focused group discussions was as follows:

(i) Students only;

(ii) Faculty members only;

(iii) Non-faculty members only;

(iv) A mixture of students, faculty and non-faculty members.

Each of the focused group discussion lasted approximately two hours.

Overall, 61 scale items were finally generated from the literature and interviews.

These items were submitted to a panel of expert judges (marketing professors) in order to assess its content validity. The panel of experts checked the scale items for ambiguity, clarity, triviality, sensible construction and redundancy, as well as to make sure that the items reflected the definition of Trust. After the elimination of 27 redundant

items or “not representative” items, the experts agreed that the scale items of Trust adequately represented the construct. The revised Trust scale had 34 items ranging from 1 “strongly disagree” to 5 “strongly agree.”

B. First study

1. Sample and data collection

The unit of analysis in this study is the individual consumers, who either has never purchased an item online or those who have purchased some items in a varied period of time.

Early data collection for item refinement was undertaken with members of the community of a university in Kenya (the survey was conducted by Google forms).

The Google forms described the research purpose and invited each receiver to participate in the survey by filling in the attached e-questionnaire. Online possesses numerous advantages over conventional interviewing methods. Online surveys offer a more efficient and convenient form of data collection [1]. In addition, an online approach can be more effective for identifying and reaching online shoppers.

IV. PRELIMINARY RESULTS

Preliminary results show that of the respondents so far, they are aged between 18 -55 years, 56% male and 44 percent female, ranging between diploma students to PhD holders. We realize that 50% of them buy at least 2 items online per month, 25% buy 3-5 items online every month, 6% buy more than 10 items online while 19% do not shop online. We also realize that majority (38.6%) spend only between KES 1 and KES 10000, as shown in Fig. 3.

Fig. 3. Average Amount spent in Kenyan shillings per month.

As can be seen, this is paltry as compared to the number of items that is potentially being bought from the brick and mortar shops.

We then interrogate our results to ascertain the contributing factors that have led to that low uptake of online shopping and we were able to confirm that trust issues are a major impediment to the uptake of online shopping in Kenya. To be precise, we realize that of the key factors that prevent them from shopping online, the leading one is fear of deception (40.7%) followed by unreliability of online services (22.8 %) then online security site (19.5%) and privacy (12.9%) as shown on Fig. 4.

Fig. 4. Factors that prevent Kenyans from shopping online.

We were also able to confirm that the participants appreciate the benefits of online shopping, consistent with literature, as shown in Fig. 5 below.

Fig. 5. Factors that encourage Kenyans to shop online.

43.2% agree that online shopping saves time, 35.7% feel that if done right can be more reliable since they can track their purchases as it gets delivered while 26.1% agree that it is cheaper than brick and mortar shops, 12.4% say they buy impulsively because of disruptive online marketing which they would not have been achieved be it not for online marketing and again 20% say they do not shop online at all.

A. Exploratory Factor and Item Analyses

We used R statistical software, with psych and GPArotation rotation packages to carry out Exploratory Factor Analysis. Parallel analysis was done to determine the number of factors for this dataset with the use of the parallel function of the psych package’s fa object (fa.parallel). We use Minimum residual (minres) as the factor method. The suggested number of factors is 5.

Fig. 6. Parallel Analysis Cree Plot.

1) Factor Analysis

After getting the probable number of factors, we then used psych packages fa() function to perform the EFA, passing the following arguments:

- r: Raw data (can sometimes be the correlation or covariance matrix).
- nfactors – Number of factors to extract (5 in our case).
- rotate – Although there are various types of rotations, `Varimax` and `Oblimin` are the most popular.
- fm – One of the factor extraction techniques like `Minimum Residual (OLS)`, `Maximum Likelihood`, `Principal Axis` etc.

We therefore have the following R code:

```
Factors<-fa(data,nfactors=5,rotate=
"oblimin",fm="minres")
```

Loadings:	MR1	MR3	MR6	MR2	MR5	MR4	MR7	MR8
gender								
age						0.661		
marital.status					0.480	0.310	-0.306	
education					0.745			
employment								
occupation								
income						0.393		
hours.online.perweek				0.361			0.717	
amount.per.purchase				0.916				
average.per.item				0.895				
amount.per.month				0.888				
preferred.shop								0.421
easy.to.understand	0.848							
terms.conditions.displayed	0.928							
company.information.displayed	0.753							
offer.secure.payment	0.698							
confirm.transaction.details	0.657							
adequate.l.features	0.764							
explain.information.use	0.461							
necessary.personal.information	0.483							
X2.policy	0.323			0.398				
exaggerate.benefits		0.620						
not.entirely.truthful		0.673						
misleading		0.704						
take.advantage		0.863						
not.needed.items		0.836						
price.shown.billed				0.708				
get.what.you.ordered				0.765				
products.available				0.710				
time				0.838				
SS loadings	4.340	2.950	2.856	2.678	0.983	0.854	0.848	0.443
Proportion Var	0.145	0.098	0.095	0.089	0.033	0.028	0.028	0.015
Cumulative Var	0.145	0.243	0.338	0.427	0.460	0.489	0.517	0.532

Fig. 7. Factor loadings.

2) Adequacy Test

After achieving a simple structure as in Fig. 7, we validated our model for adequacy as per Fig. 8.

```
MR7 0.21 0.03 0.20 0.27 0.10 -0.21 1.00 0.04
MR8 0.20 0.20 0.20 0.09 0.07 0.14 0.04 1.00

Mean item complexity = 1.7
Test of the hypothesis that 8 factors are sufficient.

The degrees of freedom for the null model are 435 and the objective function was
The degrees of freedom for the model are 223 and the objective function was 1.74

The root mean square of the residuals (RMSR) is 0.02
The df corrected root mean square of the residuals is 0.03

The harmonic number of observations is 241 with the empirical chi square 84.4 w
The total number of observations was 241 with Likelihood Chi Square = 389.01 w

Tucker Lewis Index of factoring reliability = 0.932
RMSEA index = 0.055 and the 90 % confidence intervals are 0.046 0.065
BIC = -834.09
Fit based upon off diagonal values = 1
Measures of factor score adequacy

Correlation of (regression) scores with factors MR1 MR3 MR6 MR2 MR5 MR4
0.97 0.96 0.96 0.97 0.86 0.83
Multiple R square of scores with factors 0.95 0.92 0.93 0.95 0.74 0.69
Minimum correlation of possible factor scores 0.90 0.84 0.86 0.89 0.48 0.38

Correlation of (regression) scores with factors MR7 MR8
0.85 0.73
Multiple R square of scores with factors 0.72 0.53
Minimum correlation of possible factor scores 0.44 0.06
```

Fig. 8. Adequacy Test.

The root means square of residuals (RMSR) is 0.02. This is acceptable as this value should be closer to 0. Next, we should check the RMSEA (root mean square error of approximation) index. Its value, 0.055 shows the good model fit as it's below 0.06. Finally, the Tucker-Lewis Index (TLI) is 0.93 – an acceptable value considering it's over 0.9.

3) Naming the Factors

After establishing the adequacy of the factors, we name the factors as follows:

(i) MR1 variables were named Security.

These included:

Site's Security Policy easy to understand.

Site's Terms and Conditions are displayed.

Site owner's background information displayed.

The site offers secure payment methods.

You can confirm the details of the transaction before paying.

The site has adequate security features.

The site clearly explains how user information is used.

Only the personal information necessary for the transaction to be completed needs to be provided.

Information regarding the privacy policy is clearly presented.

(ii) MR3 variables were named Trust.

These included:

The site exaggerates the benefits and characteristics of its offerings.

The site is not entirely truthful about its offerings.

The site uses misleading tactics to convince consumers to buy its products.

This site takes advantage of less experienced consumers to make them purchase.

This site attempts to persuade you to buy things that you do not need.

(iii) MR6 variables were named Reliability.

These included:

The price shown on the site is the actual amount billed.

You get what you ordered from this site.

The products I looked at were available.

Promises to do something by a certain time, they do it.

(iv) MR2 variables were named Purchasing power.

These included:

Average monthly Income.

Average amount spent per purchase.

Average amount spent per item.

Average amount spent per month.

(v) MR5 variables were named Exposure to information.

These included.

Level of Education.

Employment Status.

(vi) MR4 variable which was only one is number of hours an individual spends online per week.

(vii) MR8 variable which was also only one and is the shop preference. Some people just prefer a specific shop, due to factors not captured by this study.

V. CONCLUSION, RECOMMENDATION, AND FUTURE WORK

From the results, we have seen that Kenyans are spending very little amount of money online on average, with a majority of them (38.6%) spending between KES 1 and KES 10,000 which is paltry, as compared to what they spend brick and mortar shops. The key impediment to this is trust issues as the majority of the participants (40.7%) have shown that they fear deception while 22.8% are worried about the unreliability of online services, 19.5% are worried about online security and another 12.9% are worried about privacy, on the other hand, the participants have again shown through results that they are cognizant of and appreciate the benefits of online shopping, consistent with

literature. It is therefore recommended that the impeding factor which is trust needs to be worked on in order to help the shoppers in Kenya maximize from the benefits of online shopping.

From literature, we have also been able to confirm that when trust is incorporated into a recommender system algorithm, the performance, in terms of accuracy improves, however the model to construct this trust for computational purpose is still missing in literature since previous studies used explicit expression from human beings about whether they trust a particular product or not. This is not only a reactive approach, but also requires too much effort from online shoppers, which many at times they are not willing to part with and therefore an unreliable approach. We have therefore tried to solve the problem by studying and modeling trust from its indicators for computational purposes. Even though this study is part of ongoing work, preliminary results have shown the factors that shoppers in Kenya consider when making online purchases include:

- (i) Security;
- (ii) Trust;
- (iii) Reliability of the site;
- (iv) Individual's purchasing power;
- (v) Individual's Exposure to information;
- (vi) Number of hours an individual online spends per week;
- (vii) Specific Online Shops.

This information can be used by online shop owners as a guideline on how to improve their online platforms in order to increase revenue from their online shops.

The information can also be sensitized on what to look for in a site before making a commercial commitment to purchase from a shop or not.

Moving into the future work, from this Exploratory Factor Analysis, we are now going to do a nationwide survey for the whole country and use the data from the nationwide survey to do Confirmatory Factor Analysis (CFA Tests).

If the CFA test is OK, we will then do Structural Equation Modeling (SEM) analysis to study the Structural Relationship between the factors that contribute to trust from the eyes of the Kenyan Shoppers.

The structural relationship will give us an equation which can be programmed for computational use, and help us in incorporate some sort of trust adjustment factor to the output of a standard recommender system algorithm, automatically in order to

1) Filter out fraudulent services in a context-aware fashion in order to cover online shoppers from fraud.

2) Improve the accuracy of recommender algorithm in order to maximize on the limited time a shopper is online and willing to view output of a recommender system in order to make a purchase, thereby increasing revenue to the online shop owner.

REFERENCES

- [1] Beth, T., Borchering, M. and Klein, B. (1994) 'Valuation of Trust in Open Networks', in *Lecture Notes in Computer Science*, Karlsruhe: European Institute of System Security.
- [2] Ding, C., Yueguo, C. and Weiwei, C. (2005) 'A Survey Study on Trust Management in P2P Systems', *National University of Singapore*.
- [3] Hwang, C.-S. and Chen, Y.-P. (2006) 'Using Trust in Collaborative Filtering Recommendation', *Department of Information Management, Chinese Culture University*, January.
- [4] IEEE (2011) *Bandwidth Trends on the Internet. A Cable Data Vendor's Perspective*, 1 September, [Online], Available: HYPERLINK http://www.ieee802.org/3/ad_hoc/bwa/public/sep11/cloonan_01a_0911.pdf http://www.ieee802.org/3/ad_hoc/bwa/public/sep11/cloonan_01a_0911.pdf [30 November 2014].
- [5] Jiang, C., Liu, S., Lin, Z., Zhao, G., Duan, R. and Liang, K. (2016) 'Domain-aware trust network extraction for trust propagation in large-scale heterogeneous trust networks', *Knowledge-Based Systems*, August.
- [6] Justwana, F., Bakker, R. and Berejikian, J.D. (2017) 'Measuring social trust and trusting the measure', *The Social Science Journal*.
- [7] Koren, Y. and R. B. (2011) 'Advances in Collaborative Filtering', in Ricci, F. and Rokach, L. *Recommender Systems Handbook*, New York: Springer.
- [8] Lu, J., Wu, D.W., Mao, M., Wang, W. and Zhang, G. (2014) 'Recommender System Application Developments: A Survey', *Decision Systems & e-Service Intelligence Lab, Centre for Quantum Computation & Intelligent Systems*, January.
- [9] Park, H.J. and Kang, J. (2017) 'A Secure-Coding and Vulnerability Check System Based on Smart-Fuzzing and Exploit', *Neurocomputing*, November.
- [10] Ricci, F., Rokach, L. and Shapira, B. (2011) *Introduction to Recommender Systems Handbook*, New York: Springer Science+Business Media, LLC.
- [11] Sergio, R. (2007) 'The Ethics of Online Retailing: A Scale Development and Validation from the Consumers' Perspective', *Journal of Business Ethics*, no. 72, pp. 131-148.
- [12] Shani, G. and Gunawardana, A. (2011) 'Evaluating Recommendation Systems', in Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B. (ed.) *Recommender Systems handbook*, New York: Springer.
- [13] Victor, P., Cock, M. and Cornelis, C. (2011) 'Trust and Recommendations', in Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B. (ed.) *Recommender Systems Handbook*, New York: Springer Science+Business Media, LLC 2011.
- [14] Victor, P., Cornelis, C., De Cock, M. and Teredesai, A.M. (2009) 'Trust- and distrust-based recommendations for controversial reviews', *IEEE Intelligent Systems*.
- [15] Wang, Y., Cai, Z., Yin, G., Gao, Y., Tong, X. and Han, Q. (2016) 'A game theory-based trust measurement model for social networks', *Compu Social Networks*.
- [16] Welch, M.R., Rivera, R.E.N., Conway, B.P., Yonkoski, J., Lupton, P.M. and Giancola, R. (2005) 'Determinants and Consequences of Social Trust', *Sociological Inquiry*, vol. 75, no. 4, October, pp. 453-473.
- [17] White, T. (2012) 'Hadoop: The Definitive Guide', in Blanchette, M.L.a.M. (ed.) *Hadoop: The Definitive Guide*, Third Edition edition, California: O'Reilly Media, Inc.
- [18] Yin, C., Wang, J. and Park, H.J. (2017) 'An Improved Recommendation Algorithm for Big data Cloud Service based on the Trust in Sociology', *Neurocomputing*, July.
- [19] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [20] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [21] K. Elissa, "Title of paper if known," unpublished.
- [22] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [23] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [24] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.

Edwin O Ngwawe is a Kenyan Nationality. He holds a Master of Science Degree in Distributed Computing Technology from University of Nairobi. He is currently a lecturer at the Technical University of Kenya. He is also an active member of the institute of Electrical and Electronics Engineers (IEEE).

Elisha O. Abade is a Kenyan Nationality. He holds a Doctor of Engineering in Computational Science and Engineering and Master of Engineering in Electrical Engineering and Computer Science from Nagoya University, Japan. He is also a Certified Information Systems Auditor.

Stephen N. Mburu is a Kenyan Nationality. He holds a PhD In Information Systems (Health Informatics) and Master of Science In Information Systems from University of Nairobi, Kenya. He is also an active member of the institute of Electrical and Electronics Engineers (IEEE).