Digital Taxation: An analytics study of data mining and pattern recognition to enhance tax compliance on online business transaction

Jedidah Wanjagi[1] Kipkirui Ruto[2]

[1] Department of economics, Finance and Accounting, JKUAT University, Nairobi Kenya
[1] Department of Finance and Accounting, KAG University, Nairobi Kenya

E-mail: jedidahagnes@gmail.com

Abstract

This paper sought to determine how tax administrators could use data mining and pattern recognition to enhance tax compliance on online business transactions. The specific objectives of the study were to examine technology required in adopting data mining; to determine the tax audit and control required to detect error and fraud in data mining; and to determine the risks involved in data mining and pattern recognition to enhance tax compliance on online business transactions. The latent role and benefits of data mining in tax administrations are elucidated in view of the overall technology, operational framework and organization. The researcher reviewed various articles, research papers and books on various data mining applications. Techniques used for data mining included statistical techniques, decision tree and neuro network technique. Findings indicate that decision tree and neural network technique provided better results than the other techniques. The predictive modeling using the "Delphi" method was discovered as perfect tool that assisted agency to differentiate non-compliance from compliant clients and to focus on audits that would lead to a positive tax adjustment. The KRA may consider the use of this model to predict the risk involved in data mining. This actually assists the tax authorities to make better use of human personnel and therefore minimize the tax burden. The process of data mining helps the tax administrators to refine its traditional audit strategies in order to raise their tax budget

Key Words: Data mining, Pattern recognition, Tax Compliance

1. Introduction

Technology and the use of internet to process large amount of data concerning an increasing number of taxpayers has become a very common element in today's world (Chiang & Limato, 2017). Tax administrators need this data in order to establish the right amount of taxes for the taxpayers. In many countries across the world, taxpayers and third parties are required to provide relevant tax authority with data for the purposes of taxation (Braithwaite, 2017). As such, it is necessary that taxpayers have enough knowledge on how to comply with the requirements of the tax laws given that it's a compulsory exercise. Accordingly, tax administrators play a fundamental role in ensuring that high tax compliance level is achieved among the taxpayers (Kira, 2017).

Kenya is one of the countries which is considered to be digitalized but information poor (Rothwell & Mwachinga, 2018). This is because most of the businesses conduct their daily transactional operations through the use of internet and technology. This makes available huge amount of data that can be generated inform of texts, figures, audio, videos, and hypertexts format through the application of data mining techniques and thus using it to validate tax compliance for taxpayers (Kira, 2017). Data mining basically is the process of extracting hidden information from a wider organization's databases (García, Luengo & Herrera, 2015). This technique has previously been used by scientists, businesses and government to sift through large volumes of data in the passenger records, airline business and supermarkets scanner data (Shu et al., 2017). This technique, therefore, would significantly assist the taxing authority in generating huge amounts of data which would enhance tax administration.

2. Problem Statement

Kenya Revenue Authority is currently facing tax administration challenges due to a high tax avoidance and evasion particularly from digital organizations (Olbert & Spengel, 2017). This is due to lack of enough information to the tax administrators to equip themselves with traditional methods that provide loopholes for the tax evaders to take advantage. This practice has continuously resulted in a substantial loss of revenues to the government. Ndemo & Weiss (2016), pointed out that the tax authority should create and facilitate an enabling environment through an enactment of relevant laws and regulations that would aid in tax compliance. Dong & Wu (2015) argued that most of the organizations considerably use customized online business technology in their day-to-day processing of transactions. In addition, they also conduct simultaneous processing and storage of an equivalent data in their organizational' repositories (Assante et al., 2016). This paper, therefore, was motivated with the belief that there exists large amount of data available in the organization's databases which have remained under-utilized in discovering and supporting tax compliance

3. Research Objectives

i.   To examine technology required in adopting data mining to enhance tax compliance on online business transactions

ii.   To determine the tax audit and control required to detect error and fraud in data mining to enhance tax compliance on online business transactions.

iii. To determine the risks involved in data mining and pattern recognition to enhance tax compliance on online business transactions

4. Literature Review

Business systems which use data mining need to be supported with appropriate infrastructure and tools that aid in the processing and accumulation of large amounts of data set besides, suitable delivery and deployment of the mining results (Almunia & Rodriguez, 2018). Most of the organizations uses a customized online business technology in their day-to-day processing of transactions (Lee et al., 2019). As such, they conduct simultaneous processing and storage of an equivalent data in their organizational' repositories. Accordingly, these repositories could create, in principle, a fundamental source of input data for mining, although, it remained difficult to extract data from them directly due to some risks and problems involved (Ming et al., 2018). These comprise of quality, coverage and structure of the data stored in their databases and thus might not be up to the standard needed for mining. Other include complex queries which are regularly needed for data mining hence may lower their operational tasks performance (Eldén, 2019).

Eldén (2019) observed that the use of data warehousing would remain imperative in solving the problem of data mining where a central repository of data is regularly extracted, transformed and loaded from various selected sources. These sources might comprise of either external or internal operational databases and hence maintaining consistency, integrity and appropriateness of the uploaded data to be use for its intended objective such as analytics (Fan, Xiao, Li & Wang, 2018). However, the demand for distinct data warehouses could remain diminishing given that dealers are now enhancing their operational databases in order to improve their business systems for analytical complex queries. Therefore, a separation of analytical information systems and online business transactions might significantly decrease.

Abdel-Basset et al., (2019) argued that there exist numerous potentials of data mining which enhance a selection

of organizations for tax audit and control in today's constant increased in the complexity of business associations. The manner through which tax auditors have acknowledged the tax audit risk and control to date depends heavily on the previous knowledge to pose complex queries into the financial reports databases and thus may pick a less qualified individual for tax audit given the existence of varieties of factors for encourages tax evasion (Ordóñez & Hallo, 2019, April).

As No, Lee, Huang & Li (2018) noted, the current audit selection technique focuses primarily on a single company at any given particular time but then again there is an upsurge need of considering several organizations simultaneously. Ordóñez & Hallo (2019, April) in their study developed a wide model that helped in identifying companies that give merit to a tax audit. They further posit that these companies would be use as a target group. Subsequently, data mining would then be applied in defining the profiles of such companies which have been selected for audit tax reasons in situations where audits would have developed additional taxes (Tian & Gu, 2019). Additionally, Eldén (2019) found out that some of the characteristics that distinguish profits generated from these organizations did not require any audit and thus remain interesting.

Self-organizing network based on clustering is use as a tool for data mining so as to find similarities of the audited companies of the target group (Tian & Gu, 2019). Therefore, they act like a neutral map frequently useful when carrying out data mining work. Moreover, self-organizing network mapped data into two -dimension and thus putting observations into smaller groups. As a result, it would then integrate both clustering and visualization of the data mining tasks (Tucker, 2018). The primary aim was to establish a self-organizing network that would form a key cluster and thus would considerably accommodate majority of its target group of companies.

Tax compliance risk management deals basically with an optimization of allocated resources to a tax administration with a view of ensuring maximum achievement of compliance in accordance with the tax laws (Groulx & McGregor, 2018, June). For tax compliance, there are four categories identified for taxpayer's compliance and they entail registration in the system, timely filling of returns, reporting accurate information and making timely payment of tax obligations (Alm, 2019). The tax compliance method takes into consideration the diverse capabilities, attitudes, behaviors and motives of the taxpayers and thus there is a demand to making adjustment of the tax administration's service.

Siglé et al., (2018) found that tax compliance risk comprises of taxpayer's failure to accurately register, accurately file their tax information, accurately report liabilities and accurately make payment of the tax obligations. Fauziati & Kassim, (2018) provides a general method which defines risk as a probability that an event would unfavorably influence an ability of the organization to realizing its goals. Tax compliance method has become a center of risk management and thus supporting in the process of identification, analysis and reduction of tax risks. According to Fauziati & Kassim, (2018), tax audit risk management is the process where tax administrator makes thoughtful choices on how treatment tools could be applied to encourage compliance and averting non-compliance, depending on the knowledge and behavior of the taxpayers besides their available capacity (Siglé et al., 2018). Moreover, the fundamental of applying risk compliance management is to facilitate tax administration in order to establish its mission(s) through improving management and thus better decisions making.

5. Research Design and Methodology

The researcher reviewed various articles, research papers and books on various data mining applications.

5.1 Data Mining

Scholars argue that advanced analytics would play a big role in tax administration therefore enhancing the level of compliance. An advanced analytical study should support taxpayers through tailoring services according to the relevant needs and tax issues of a large group of taxpayers. Services designed to cover specific tax issues enhance the level of tax compliance from specific taxpayers as there able to understand and feel less burdened with the relevant tax laws in place.

Analytics is important as it makes taxpayers feel appreciated. This is through the equal treatment of taxpayers. It is used to analyze, identify and indicate incidences where taxpayers perform registration and filing of reports incorrectly. Enhanced analytical study ensures detection of such events so that taxpayers pay the lawful amount required by the tax authorities.

There basically two broad groups of data mining tasks; the descriptive task and the predictive task. According to Gupta (2012), descriptive task actually includes cluster analysis, association of rules and describing data usage. Descriptive task is quite important as it is used to create a model of different behaviors and transactions which are quite suspicious.

Predictive task is slightly different as it makes a forecast for each particular observation. It utilizes regression analysis to determine the relationship between one or more variables. The statistical technique for these model includes; the linear regression, decision trees and neural networks (Thillainayagam, 2012).

5.2 Data Mining Algorithm

Data mining algorithm is a well devised set that inputs data and produces output in form of models and patterns (Hand et al., 2001). The process is made up of finite set of rules and procedures.

5.3 Technology Required in adopting Data Mining to enhance Tax Compliance on Online Business Transactions

Mining Methodology

Most researchers tried developing new different data mining technologies. The technology required an audit depends on different levels of knowledge mining in a specified space, level of integration and consideration of different semantic ties amidst various data objects. Some mining techniques also explore specific measures that can be applied for the discovery of relevant patterns to assist in the relevant discovery process. These mainly adapted techniques include:

Statistical Technique

In this method, predictions are done using a synonymous regression model. A model is created mapping the values from the predictors in a way that error is minimized when analyzing. This method was important as it enhanced the tax audit and control mechanisms. Simplest form of regression was the linear regression model (Thai-nghe, 2010).

Decision Tree Technique

In a decision tree each branch is a classified question and therefore the leaves become the partitions in the dataset with their own classification (Sheard & Hurst, 2003). Objects are categorized by following a path down the tree. Decision tree are quite important in estimation and classification task. It can be used to preprocess data for different algorithms. In data mining, a decision tree describes data but not decisions and the resulting classification tree can be input for decision making.

Neural Network Technique

A neural network is defined as a collection of different interconnected neurons (Guo & Zhang,2009).The neural network comprises two main structures; the node which closely corresponds to the neuron found in the human brain and the link which correspond to the lines connecting the human brain (Blagojevic, 2013). The arrangement of the neurons is termed as an architecture of the network. Depending on the type of the neural network technique, they correspond to different architectures. Different types of neural networks are used for discovering the weakness and strengths of the interconnections.

Neural networks encompass sets of connected input/output units, where each has a different connected weight. During the "learning phase'' the specific weights are adjusted on the neurons to allow the network to extract patterns that can actually be used for prediction (Thillainayagam, 2012).

Decision tree and neuro network technique would be quite relevant for this research as they both involve in-depth investigation. The neuro network would be used for clustering, classification, prediction and pattern recognition while the decision tree results from classification can actually be used for decision making. Ranjan & Khalil (2008), applied the decision tree in data mining and found out that it was relevant for decision making purposes. Blogojevic (2013), applied both the neuro network and the decision tree in data mining and discovered that these two techniques were useful in data mining for companies and the self-employed personnel. KRA can use these two techniques when gathering data on the tax payers

5.4 Data Mining Process

The Cross-Industry Standard Process for Data Mining (CRISP-DM), was developed by Chrysler and NCR (1996), who were data mining consultants using real-life experiences to build a six phase data mining process incorporating basic knowledge and different goals of organizations

Figure 2: CRISP-DM Process Model
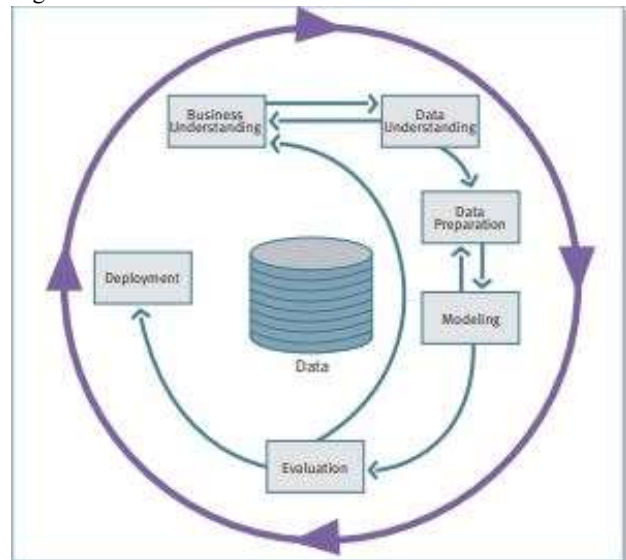
Source: IBM (2011)

1. Business Understanding

The first step ensures all participants are well conversant with the goals of the project from an organizational or business perspective. The goals of the relevant tax authority are incorporated into the data mining delinquent and a detailed type of a project plan. For tax authorities, this involves a consensual understanding of the audit process, functions and roles performed, level of evidence gathered and the prevailing challenges while trying to improve the efficiency of the audit process (IBM,2011).

2. Data Understanding

This phase involves access to quality sources and specific characters present in the data. Information gathered from the first phase would act as a cornerstone in providing more insight into the project. The specific results became a key element of the relevant dataset that would be used to generate the models (IBM,2011).

3. Data preparation

This involves placing the data into formats well suitable for the model. The researchers use the business objective to

determine the best data types and algorithms to incorporate. It also involves data issues that were not covered in the previous phase including data that is missing(IBM,2011).

    4.   Modelling

This phase involves algorithm building which is then used to extract data from the knowledge generated. There various data mining tools that are relevant for identifying a dataset. The KRA would then use classification or a regression model. The phase produces knowledge model in an appropriate format (IBM,2011).

    5.   Evaluation

The evaluation phase deals with evaluation of the models' quality. Data mining algorithms that uncover the unlimited number of patterns. It assists in determining which kind of models are quite useful in achieving the relevant business objective(IBM,2011).

    6.   Deployment

During this phase, an organization uses the results from data mining in normal business activities during decision making process. A major or minor modification may be required during decision making depending on the significance of results. The phase might involve a continuous process of model enhancement. Different tax laws and procedures may change therefore, analyst require a suitable process for regularly updating the model and bringing new outcomes (IBM,2011).

The six phases described above are quite important in every data mining process. Although the phases are not static and may change over time depending on the data needed. The model is built from various techniques and tools. The data mining techniques like the ones described above come from various disciplines including statistical analysis, machine learning, signal processing, pattern recognition and pattern visualization (IBM, 2011). For a typical data mining most of the time is spent on the model rather than the work performed in this phases.

5.5 Case Application of Data Mining Technique

In order to test the different data mining techniques that suite specific countries and different situations, the researchers looked at different countries that were practicing data mining in their respective countries.

The table below indicates the various data mining techniques that have been applied by several countries to detect online tax evaders.

Technique Applied	USA	Canada	Australia	UK	Bulgaria	Brazil	Peru	Chile
Neural Networks	✓	✓		✓	✓		✓	✓
Decision tree	✓	✓	✓				✓	✓
Logistic Regression	✓		✓	✓	✓			
SQM			✓					✓
K-means			✓					✓
Support vector machines	✓		✓					✓
Visualization Techniques	✓					✓		
Bayesian Networks			✓					
K-Nearest Neighbour			✓					
Association Rules							✓	
Fuzy rules							✓	
Markov chains						✓		
Time series		✓						
Regression				✓				
Simulations	✓							

From the figure, different techniques have been used to detect the tax evasion from online business transaction. Tax administrators can combine different formats depending on the type of business transaction. USA applied neuro networks, decision tree, logistic regression, support vector machines, visualization technique and simulations. Further research indicates that there were successful in detecting the tax evaders and avoiders who were forced to pay the taxes or force closure of their businesses. Canada has applied three different techniques which include neuro networks, decision tree and time series. Australia has applied decision tree, logistic regression, SQM, K-means, support vector machines, Bayesian networks and K-nearest neighbor. UK has applied neural networks and logistic regression, Bulgaria has also applied neural networks and logistic regression, Brazil has used visualization techniques and Markov chains, Peru used neuro networks, decision tree, association rules and fuzy rules and lastly Chile used neural networks, decision tree, SQM and K-means. From the table the mostly used is the neural networks and the decision tree. This is so because of their ability to capture data by using neurons for the case of neural networks. For the case of the decision tree we use the available data to link us to the next data until finally we get data that can be used to show the tax evaders and avoiders.

United Kingdom Use of Data Mining

HMRC discovered that they could actually use neuro-networks or big data to stop tax evasion and avoidance. The strategy employed was data mining of social media content to indicate the people that lived a better life than the one which is evident in the taxes they pay. According to the tax administrators the social media was the reason they received

losses in taxes every year of over 247.49 GBP each year (Rohrlich,2018).

Tax evasion and avoidance was costing approximately 5,357,850 GBP each year (Rohrlich, 2018). Large companies were evading taxes and the tax administrators decided to develop an analytics study to drill data from its citizens. The tax administrators decided to increase the amount of data HMRC can hold and analyze (Reeds, 2015). They were able to do this through extending their legal rights to gather data from specific online trading business including those based outside UK. This was a very crucial step since tax avoidance schemes had based their business overseas. They also extended their power to administer online businesses that sells goods without paying taxes for it. The second phase was data gathering in order to monitor the different groups that existed. These included the tax evaders and avoiders, whom data was gathered from and they were later convicted for the tax crime committed. Majority of the online business people that had oversees business were also closely monitored as they have a higher chance of evading taxes. The final phase was changing the behavior HMRC by data usage from reactive to proactive (Reeds, 2015). They used insights into human behavior by pushing them to file taxes and be very honest in the filing process.

5.6 Models in Data Mining Technique
Predictive Modelling

This is the most commonly used data mining technique. It assists organizations to actually predict relevant results of the process and use it to derive an expected outcome. In an audit selection, main target is to predict the best audit that will yield to greater tax achievements (Sumiran and Keerthi, 2018).

Predictive model uses a numerical score to indicate the best results in an audit process. A score that is high signifies the tax adjustment in the audit is higher above the average, besides a score that is low would indicate a significantly low probability of a large modification of the tax. A perfect example is the Discriminant Index Function (DIF) used to indicate returns having a higher probability of not declaring their outcome. Using DIF score would result in significant higher tax collection than random audits.

Risk Involved in Data Mining to Enhance Tax Compliance on Online Business Transactions
Delphi Prediction of Risk

1	2	3	4	5
Very high	High	Medium	Low	Very Low

According to this model introduced by the Federal Public Service of Finance in 2001, it was able to predict the risk involved during collection and recovery administration (Hondt, 2001). Delphi technique was used to predict the risk

involved. It had 3 version one for self-employed, one for all the companies and another for production. In order to minimize work involved in operational sector, the number was limited to five with each having a different color. Red to indicate the high risk involved and at the end green to represent the low risk involved.

They further went ahead to predict payment behaving using Delphi red score using 'HERMES'. 'HERMES' was able to predict if an individual or a company could be able to pay taxes in a period of 1 year (Hondt, 2001). The population that was used to come up with the model were all physical persons or companies that had a red color (high risk) in the Delphi model.

Hondt (2001), went ahead and introduced 'IRIS' to predict future payment resulting from follow up of phone calls. 'IRIS' indicated whether or not a person or company will pay their taxes within 14 days after having received a phone call from the Debt Relationship Management unit. The same can be used for Value Added Tax for self-employed and companies and the scores would indicate the level at which one would comply or not comply to paying taxes after receiving phone calls. It could also be applied to withholding tax on income where the companies are the only that would be investigated.

This model would be the best model to be used by the Kenya Revenue Authority to predict the risk involved in data mining. The KRA would be able to tell which businesses are paying taxes and the probability of them doing so after follow up on phone calls. It would be relevant as the cost of operation is minimized since the model itself does the prediction and therefore less workers are involved.

5.7 Tax Audit and Control required to detect error and fraud in data mining to enhance tax compliance on online business transactions

Advanced database system (ads) usage to determine outcome of sales tax audit on online businesses

The ADS is a division data of an audit designed warehouse to assist in applications of tax compliance like an Audit Select system, which uses the predictive models in order to identify audit of regional sales tax.

Figure 1: Model Training and Scoring
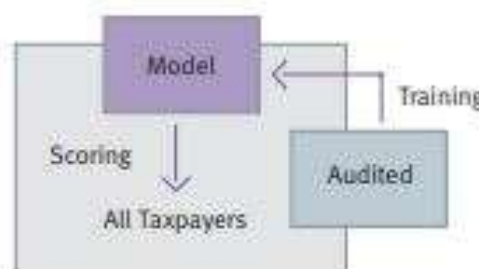
During the first step the model is calibrated using a specified training set containing a historical audit having a familiar outcome. This would enable the model to learn different relationship about the taxpayer attribute and actual audit outcome. The model uses specifically five types of data in order to generate a profile:

1.   Information from the business including taxpayer SIC code, the type of business and location
2.   Sales tax filings using recent sales reports of taxes
3.   Any type of tax filing done just for franchise tax
4.   Information of wages from employees that have been reported to any other state agency
5.   Past audit results

Figure 2: Data source for the audit selection model.

Model Inputs
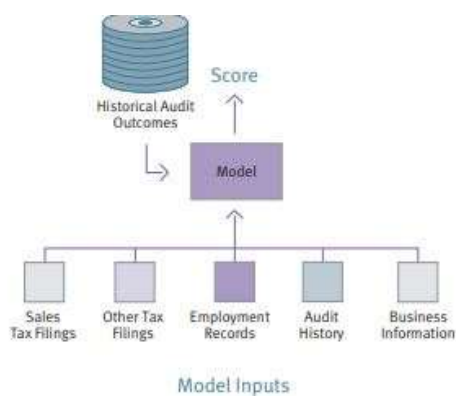
After the model is trained it can then be used for all taxpayers through a process termed as generalization. The model would generalize its learning simply from the historical audits and then analyze the data returns and finally assign the right score for the audit. The human auditors then can use through audit select score to distinguish which online business to start auditing.

Training set of Historical Audit

Gross Sales	SIC Group	Wages	Receipts	Tax Adjustment
$21,11,288	23	$34,456,345	$988,945	$100,202
$34,234,334	43	$11,476,544	$2,545,251	$434,323
$9,874,556	23	$45,443,343	$4,534,521	$0

Potential Audits

Gross Sales	SIC Group	Wages	Receipts	Predictive Score
$424,454,762	43	$5,454,362	$22,571,243	760
$44,572,462	32	$45,445,623	$45,653,235	450

Source: IBM, 2011

According to the figure, the first table indicates the historical audits of the training set. The second table indicates the predictive scores for the current population of taxpayers. According to the results, comparison is made from the actual audits and historical audits in order to discover the appropriate businesses to start auditing. The scores that match or are close to the actual outcome would be the first ones to be physically audited.

The KRA could use the ADS model when auditing the online business that fail to pay taxes. This data would be quite important to the tax authorities as they could reduce workload of doing an audit of all online businesses. The model would reduce the scope to only those having scores to undergo the audit process.

6. Discussion and Findings

The findings indicate that the Kenya Revenue Authority (KRA) can implement different techniques and models to discover the non-compliant from the compliant customers. The most commonly used techniques to determine the technology required in adopting data mining to enhance tax compliance on online business transactions are the neural networks and decision tree. According to the finding the neuro network techniques and the decision tree are mostly used in data mining due to there in depth analysis. The findings indicate that the two techniques would be quite important to the KRA when gathering data regarding tax payers doing online business. They could use the neuro network to cluster business according to the type of business involved and then use the decision tree to describe the data and hence make decisions. According to a research done by Sahu and Shrma(2016), decision tree and neuro network techniques are important tools for data mining purposes. The researchers discovered that the decision tree is useful in decision making while neuro network is useful in pattern recognition through data clustering. Data could be retrieved from Facebook, Gmail, Instagram and twitter. This data could be important to the tax authorities as they could be able to gather tax which will enable them achieve their revenue budget.

Findings related to the objective on the risks involved in data mining to enhance tax compliance indicate that the predictive model could be used to trace patterns of behavior among individual taxpayers. The Predictive model and Advanced Data System (ADS) are used to collect information from the online businesses and predict the evaders, avoiders and those that are compliant. The predictive model using Delphi technique assigns scores red indicates greater risk while green indicates less risk. The model has previously been applied in Belgium by the Federal Public Service of Finance in 2001, and it was able to predict payment behavior using 'HERMES' exclusively for Delphi red scoring (Hondt, 2001). The findings of the model indicate that even with less staff the predictive modeling and data analytics helped in recovering of debts by taxpayers.

The findings related to the tax audit and control required to detect error and fraud in data mining to enhance tax compliance on online business transactions indicate that the Advanced Database System could be used to predict sales tax. The ADS model would assist the tax authorities in Kenya to identify the right businesses to audit by identifying only those that have the right scores which can then undergo physical audits. Before the model assigns scores it first generalizes all tax payers then analyzes the historical data and assigns score only to the dataset that needs to be audited. The model has previously been used by 1BM (2011), to train set of historical audits in order to get the predictive scores. According to IBM (2011), the best way to assess a training score system is to compare the scores it produces to the actual outcomes. The goal is to compare the scores and the results from the physical audit in order to know the non-compliant and compliant tax payers.

The tax authorities can extend their rights to gather data in facebook, twitter, Instagram, gmail and yahoo. Then they could apply DIF to identify those that do not report their outcome and hence fail to pay taxes from handling their online businesses. They can do further investigation through the Advanced Database System (ADS) to predict sales tax audit outcomes for online businesses. The first phase would be to do a model training using historical outcomes of previous audits. Then the model can be applied to a large number of taxpayers through the generalization process. Afterwards they would analyze the returns and assign audit scores. Then experts trained from the KRA can then use the data to select which business to start auditing.

7. Conclusion

Digital taxation is a trending issue that would have an impact in achieving the revenue targets of our country Kenya. The KRA must take advantage of this new technologies and techniques to enhance tax compliance. To be successful, the administration officials must be at the forefront of learning on how to implement this new techniques and models. The researchers have discovered several methods that could be applied in the data mining process. From the findings, data mining has proven to be a vital tool to use to collect data from online businesses and the tax administrators can use this tool to gather information on fraud and online tax evasion to achieve the revenue collection budget. The predictive modeling assist agency to differentiate non-compliance from compliant clients and to focus on audits that would lead to a positive tax adjustment. This actually assists the tax authorities to make better use of human personnel and therefore minimize the tax burden. The process of data mining helps the tax administrators to refine its traditional audit strategies in order to achieve their target.

8. Recommendation

The tax authorities can use analytics in the audit process. For the KRA to move to the digitalization era they need to implement neuro networks and decision tree technique. This would assist in clustering, classification, prediction and pattern recognition while the decision tree results from classification can actually be used for decision making. Furthermore, they can also implement the Delphi method using predictive modeling to determine the risks involved in the audit process. For the revenue department to have an audit trail to detect error and fraud in tax administration it could use the advanced database system. This system keeps a record of history of all scores during an audit process. The scores indicate those that were paying taxes and the evaders. Using the scores, the revenue authority would be able to settle on a particular online business to do the physical audit. This would assist the KRA achieve its revenue budget that has been quite low in Kenya due to less compliance by the different online businesses. The first step would be enriching their employee's skill set and infrastructure. The new skill set would be used for statistical analysis of the online business. The tax administrators can train staffs showing interest in working with numerical analysis to use the different methods and techniques.

Reference

[1] Alm, J. (2018). What Motivates Tax Compliance? *Journal of Economic Surveys*, *33*(2), 353-388. doi:10.1111/joes.12272

[2] Al-Mosaileekh, S., Saleem, E., Abd El-Aziz, S., & Abdel-Basset, M. (2019). Horizontal Multistage Acid-Fracturing Formulates Development Strategy in Tight Tuba Carbonate Reservoir, North Kuwait. *SPE Middle East Oil and Gas Show and Conference*. doi:10.2118/195099-ms

[3] Almunia, M., & Lopez Rodriguez, D. (2014). Heterogeneous Responses to Effective Tax Enforcement: Evidence from Spanish Firms. *SSRN Electronic Journal*. doi:10.2139/ssrn.2469288

[4] Bolton, R., & Hand, D. (n.d.). Significance tests for patterns in continuous data. *Proceedings 2001 IEEE International Conference on Data Mining*. doi:10.1109/icdm.2001.989502

[5] Braithwaite, V. (2011). Tax Evasion. *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199844654.013.0016

[6] CAO, X., FAN, Z., WANG, S., YANG, X., & LIU, J. (2018). Research on Real Time Stacking Technique for Single Shot Recording. *International Geophysical Conference, Beijing, China, 24-27 April 2018*. doi:10.1190/igc2018-103

[7] Chiang, B., &Limato, J. (2017). The Use of Technology in Tax Preparation: A Closer Examination of Electronic Filing and Filing Errors. *Journal of Corporate Accounting & Finance*, *28*(3), 47-60. doi:10.1002/jcaf.22258

[8] Dong, J. Q., & Wu, W. (2015). Business value of social media technologies: Evidence from online user innovation communities. *The Journal of Strategic Information Systems*, *24*(2), 113-127. doi:10.1016/j.jsis.2015.04.003

[9] Fauziati, P., &Kassim, A. A. (2018). The effect of business characteristics on tax compliance costs. *Management Science Letters*, 353-358. doi:10.5267/j.msl.2018.4.013

[10] García, S., Luengo, J., & Herrera, F. (2015). Data Preprocessing in Data Mining. *Intelligent Systems Reference Library*. doi:10.1007/978-3-319-10247-4

[11] Groulx, A., & McGregor, C. (2018). A Social Media Tax Data Warehouse to Manage the Underground Economy. *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. doi:10.1109/hpcc/smartcity/dss.2018.00261

[12] Gupta, M., & Han, J. (2012). Approaches for Pattern Discovery Using Sequential Data Mining. *Pattern Discovery Using Sequence Data Mining*, 137-154. doi:10.4018/978-1-61350-056-9.ch008

[13] Hand, D. J. (2007). Matrix Methods in Data Mining and Pattern Recognition by Lars Eldén. *International Statistical Review*, *75*(3), 418-418. doi:10.1111/j.1751-5823.2007.00030_10.x

[14] Hassanien Mohammed Najiet al., H. M. (2018). The Effectiveness of Data Mining in Query Optimization. *International Journal of Computer Science Engineering and Information Technology Research*, *8*(2), 69-76. doi:10.24247/ijcseitrjun20188

[15] Huh, B. G. (2019). The Effect of Auditors' IS Audit on Audit Hours and Audit Fees. *korean management review*, *48*(1), 181-210. doi:10.17287/kmr.2019.48.1.181

[16] Kevinnne, G. (2016). Reviews of Mastering Digital Transformation (Hanna, 2016) and Digital Kenya (Ndemo " Weiss, 2016). *South African Journal of Information and Communication*, (18), 0. doi:10.23962/10539/21791

[17] KIRA, A. R. (2017). An Evaluation of Governmentsâ ™ Initiatives in Enhancing Small Taxpayersâ ™ Voluntary Tax Compliance in Developing Countries. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *7*(1). doi:10.6007/ijarafms/v7-i1/2687

[18] Noam, E. (2018). Inequality and the Digital Economy. *Digitized Labor*, 117-140. doi:10.1007/978-3-319-78420-5_8

[19] Sandford, C. (2019). Policies Dealing with Tax Evasion. *Underground Economies in Transition*, 87-100. doi:10.4324/9780429021138-8

[20] Shin, J., & Lee, S. (2019). A Study on the Effect of Online Activation Business Transaction Factors of Fresh Food Shopping Mall on e-Customer Relationship Quality and e-Customer Loyalty. *The East Asian Journal of Business Management*, *7*(1), 1-16. doi:10.20498/eajbe.2019.7.1.1

[21] Siglé, M., Goslinga, S., Speklé, R., Van der Hel, L., &Veldhuizen, R. (2018). Corporate tax compliance: Is a change towards trust-based tax strategies justified? *Journal of International Accounting, Auditing and Taxation*, *32*, 3-16. doi:10.1016/j.intaccaudtax.2018.06.003

[22] "Tax Compliance and Administration," in Handbook on Taxation; eds. Hildreth, W. B., Richardson, J. A., pp. 741-768. Marcel Dekker, Inc.

[23] Thillainayagam, V. (2012). Data Mining Techniques and Applications- A Review. *i-manager's Journal on Software Engineering*, *6*(3), 44-48. doi:10.26634/jse.6.3.1791

[24] Ziegel, E. R., Fayyad, U. M., Piatetski-Shapiro, G., Smyth, P., &Uthurusamy, R. (1998). Advances in Knowledge Discovery and Data Mining. Technometrics, 40(1), 83. doi:10.2307/1271414