International Journal on Perceptive and Cognitive Computing (IJPCC) Vol9, Issue 1 (2023)
https://doi.org/10.31436/ijpcc.v9i1.343

An Enhanced Predictive Analytics Model for Tax-Based Operations

[1]Zainab I. Musa*, [2]Sahalu Balarabe Junaidu, [3]Baroon Ismaeel Ahmad [4]A.F. Donfack Kana, [5]Adamu Abubakar Ibrahim

[1,2,3,4]*Department of Computer Science, Ahmadu Bello University, Zaria. Nigeria*
[5]*Department of Computer Science, International Islamic University Malaysia.*

*Corresponding author: cyberzaynab@gmail.com
(Received: 22nd July 2022; Accepted: 23th December 2022; Published on-line: 28th January 2023

Abstract**—** In order to meet its basic responsibilities of governance such as provision of infrastructure, governments world over require significant amount of funds. Consequently, citizens and businesses are required to pay certain legislated amounts as taxes and royalties. However, tax compliance and optimal revenue generation remains a major source of concern. Measures such as penalties and in the current times Data and Predictive Analytics have been devised to curb these issues. Such effective Analytics measures are absent in Bauchi State and Nigeria as a whole. Previous studies in Nigeria have done much in the area of tax compliance but have not implemented Data Analytics solutions to unearth the relationships which this study will cover. A Combined Sequential Minimal Optimisation (CSMO) model has been developed to analyse co-relation of Tax-payers, classification and predictive traits which uncovers trends on which to base overall decisions for the ultimate goal of revenue generation. Experimental validation demonstrates the advantages of CSMO in terms of classification, training time and prediction accuracy in comparison to Sequential Minimal Optimisation (SMO) and Parallel Sequential Minimal Optimisation (PSMO). CSMO recorded a Kappa Statistics measure of 0.916 which is 8% more than the SMO and 7.8% more than the PSMO; 99.74% correctly classified instances was compared to 98.28% in SMO and 98.35 in parallel SMO. Incorrectly classified instances of CSMO recorded a value of 0.25% which is better than 1.72% of SMO and 1.68% of PSMO. Training time of 223ms was recorded when compared to 378ms in SMO and 286ms in PSMO. A better value of 0.9981 for CSMO was achieved in the ROC Curve plot against 0.944 in SMO and 0.913 in PSMO. CSMO takes advantage of powerful Analytics techniques such as prediction and parallelisation in function-based classifiers to discover relationships that were initially non-existent.

Keywords: Taxes, Tax operation, Predictive Analytics, Optimisation

I. INTRODUCTION

Tax can be defined as a compulsory payment imposed by the government through its agents on income of individuals and corporate bodies as well as on goods and services [1]. Tax can be a direct or indirect taxes. Direct tax involves taxes imposed by the government on income of individuals and corporate organizations. While indirect taxes are taxes on goods and services, which are produced within a country, exported to other countries, or imported into a country.

The inability of many governments and other revenue-generating bodies around the world to increase tax collections, which is necessary for maintaining economic growth and catering to the requirements of their populations, continues to be a major cause for concern [2-3]. Powerful techniques of analytics could be used on the currently underutilised large amounts of data accumulated in the repository of tax administrations to achieve the ultimate goal of increasing revenue [4]. This is important in light of the existing difficulties of tax compliance and the ultimate goal of increasing revenue. In order to ensure that essential public services like education and infrastructure continue to be accessible, the Bauchi Internal Revenue Service (BIRS) conducts evaluations of property values, transactions, licences, and incomes. Taxes are what are collected from the public in order to fund the provision of certain public services [5]. In its current iteration, the BIRS makes use of the conventional database structure for the storage and retrieval of records. Records are kept in their own distinct databases, including Excel files, flat files, and many others. As the amount of data that needs to be manipulated grows, there is a growing demand for a tax administration system that is not only efficient but also fair and honest.

From the overall assessment of BIRS' tax administration business processes and interaction with relevant directorate heads and some of their essential staff, we were able to identify key problem areas that should be considered as an analytics system is being integrated. These problem areas are:

A: Existence of a wide tax gap between tax that should be collected and what is actually collected. A recent International Monetary Fund working paper categorized Nigeria as one with a highly ineffective tax administration and a tax gap in the range of 40% (IMF, 2011). According to the 2 BIRS Management staff interviewed, only 50% of the due tax is believed to be collected by the BIRS.

B: Inability to enforce compliance of tax payment due to incomplete, inaccurate and poor access to tax payers' data.

C: Tax fraud which involves legal source tax crimes (E.g., false declaration) and illegal source financial crimes(Since all income must be reported regardless of source, illegal dealings such as money laundering could easily be detected through evasion or filing patterns)

Extraction and application of previously undiscovered, complete data from big databases for the purpose of making critical business choices is known as "data mining" [6]. For the goal of generating conclusions, data analytics is the analysis of data for relationships that have not previously been uncovered. Unlike Data Mining, which merely analyses data in search of patterns without a specific goal in mind, this approach has a clear end goal in mind. Using current data to make predictions about unknown future events is known as predictive analytics.

II. RELATED WORK

Machine learning is a subfield of computer science that focuses on giving machines the ability to learn without being given specific instructions [7]. Building predictive models is one application of machine learning. Data is processed by ML by employing rules stored in the computer to automatically perform calculations, weighing, and other tasks. A computer programme is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. This definition applies to any situation in which the program's performance on T is being measured [8]. A technology known as "Predictive Analytics" is one that "learns" from its past experiences, or "data," to forecast the behaviour of individuals in the future so that better choices can be made [9].

By utilising the services of both the Naive Bayes classifier and the K Nearest Neighbor algorithm in a sequential fashion, Ferdousy et al. [10] are able to compensate for the shortcomings of both of these classifiers. The Naive Bayes classifier is based on statistics and must discretize the numerical attributes it uses into a large number of categories. K Nearest neighbour is a distance-based algorithm that calls for the measurement of the distance between categorical attributes. The research eliminates the need for discretization and distance measurement because it combines the two classifiers in a logical manner that eliminates the need for either of those procedures. It has been demonstrated that the combined model performs better than the Naive Bayes classifier and K Nearest Neighbors when used separately.

Kachele et al. [11] presented a demonstration that showed how the Sequential Minimal Optimisation technique could be used to train support vector machines in parallel. The dataset is partitioned into subsets, and these subsets are then distributed across multiple independent working nodes, each of which has a single instance of SMO. The problems faced by the subsets are tackled independently and in parallel, with the subsets communicating with one another in-between nodes and exchanging filtered multipliers.

There have been a lot of empirical research done in the past that consider using taxes in machine learning and predictive analytrics. In the modern landscape of taxation, Singh [12] presented a digital transformation tool that included the application of predictive analytics. In a similar vein, Junquera-Varela et al. [13] backed up that assertion and extended it to include the utilisation of customs administrations. On the other hand, Owens et al. [14] detail how the recent developments in the application of predictive analytics can be utilised in the tax administrations of African countries. On the other hand, Noor et al. [15] investigated its application in the context of Malaysia and discovered that machine learning techniques can be used to predict revenue for the federal government of Malaysia. In predictive analytics, the nodes make up a hierarchical structure, and as the process moves forward, the support vectors move up the hierarchy, with the most significant ones ending up at the very top and the least significant ones at the very bottom [16].

The top node verifies whether or not all of the Karaush-Kuhn-Tucker (KKT) conditions for the entire problem have been met [17]. This strategy has the benefit of not requiring the individual sub-optimization problems to be solved, but rather, only the top node will arrive at a global optimum. This is a significant time saver. Once a direction has been selected, a multiplier will proceed in that direction until it reaches the top node and then switch directions.

On the Bauchi Internal Revenue Service (BIRS) Taxpayer dataset, our research work presents the hybridised implementation of Sequential Minimal Optimization (SMO) and Parallel Sequential Minimal Optimization algorithm (PSMO). PSMO is more efficient with larger amounts of data, while SMO is more efficient with smaller amounts of data. The combination of the two, in a weighting ratio of 30:70, will result in taking advantage of the benefits offered by both classifiers, leading to improved accuracy in classification as well as prediction and interpretability. This is what previous research studies ignores. The system eliminates the drawbacks of over-fitting and learning pattern difficulties in Neural networks and Decision Trees in the study conducted by Gonzalez et al., improves the study conducted by Deepa et al. in terms of predictive accuracy, and is also anticipated to perform better in comparison to the Parallel SMO study conducted by Kachele et al. It makes use of real data provided by the Bauchi State Internal Revenue Service regarding taxpayers in the state of Bauchi.

III. METHODOLOGY

In this study, the application of machine learning is utilised as an essential step toward driving the prediction of the tax issues. Design, implementation, and testing are the three components that make up the methodology. The methodology is divided into these three parts..

A. Dataset

Data Sampling has been carried out which involves creation of training and validation data sets. This is described in Table 3.1. Six (6) attributes in the Individual database and seven (7) attributes in the organisation database were discovered after pre-processing. Two classes- compliant and non-compliant class have been used. Number of objects denotes number of records on the dataset.

Table 1:
Description of Dataset used

Dataset	No of Attributes after Pre-processing	Number of Classes	No of Objects
Individual Taxpayers	7	2	9419
Organisation Tax Payers (Pay As You Earn)	6	2	775

The data to be used for this research was gathered from the Bauchi Internal Revenue Service within assessment period of 5 years: 2011 – 2015 (See Compliance Matrix at Appendix I). The predictive model has been constructed to accept assessment period of 10 years and utilize patterns of individual and organisation taxpayer profiles, previous filings and payment information.

Dataset comprises10,200 (9419 individual and 775 organisation) instances as in Table 1 above and 22 attributes of taxpayer data in total. Our Model has the capability of taking 500,000 instances and 100 attributes of taxpayer data.

B. Model Building Attribute Selection

Attribute selection was integrated in the model and implemented using Java. It helps identify and remove unneeded, irrelevant and redundant attributes from data that do not contribute to the accuracy of a predictive model. Information gain method was employed by calculating the information gain for each attribute. The attributes that contribute more information will have ahigher information gain value. Relevant attributes are selected and those that do not add much information will have a lower score and can be removed from the entire dataset for better interpretability, accuracy and faster training time.

C. Model Building Attribute Selection

Cross-fold validation is a method of evaluating and validating predictive models by partitioning the original sample into training set to train the model, and a test set to evaluate it.

Cross-fold validation is employed to estimate accuracy of the model with test data set being independent of the training dataset to avoid over-fitting. 10-fold cross validation. A 10-fold partition of the dataset is created by breaking data into 10 sets of size n/10. For each of 10 experiments, nine (9) datasets are to be used for training and the remaining one (1) for testing. Repeat 10 times and take mean accuracy.

Sequential Minimal Optimisation Algorithm is an improved training algorithm which breaks down a large problem into a series of smaller problems avoiding the large matrix computation of training sets in between linear and quadratic time with a linear amount of memory in the training set size. SMO Algorithm is used as the basis of research because it is the most efficient classifier based on previous research.

The Taxpayer Dataset is split into two; The first part goes to the SMO which breaks the problem into series of small sized Quadratic Programming (QP) problems and solves the smallest possible optimisation problem analytically avoiding the time-consuming numerical QP computation. The parallel SMO partitions the other part of the training data set into smaller subsets and then simultaneously runs multiple CPU processors to deal with each of the partitioned data sets thereby reducing training time while keeping classification accuracy.

This method avoids the problem of over-fitting being faced in other techniques as it strategically selects a small number of data points to train based on support vector machines.

D. Experimental tools and Analysis

In this section, both the necessary materials and the procedure that should be followed are discussed. Java was used to accomplish everything from the Graphical User Interface (GUI) to classification, mining implementation, and multi-core programming. All of these tasks were successfully completed thanks to Java. We utilised the PHP (Hypertext Preprocessor) and MySQL database management systems, as well as the Pentaho Data Integration tool, for the purposes of data cleaning and preliminary processing. The programming language Java was selected to be used as the language for machine learning due to the following reasons: it is open source; it can be used to carry out data analytics and predictive modelling; it requires a short amount of time for training and testing; it can be used to carry out parallelization; and finally, it requires a short amount of time for training and testing. PHP and MySQL were used to write the database scripts that have been written to perform operations such as merging, splitting, and manipulating data and tables. These scripts have been written in order to perform the aforementioned operations. The Kettle tool from Pentaho Data Integration, which is also available as open source, can be used to extract data from a variety of databases, carry out any necessary transformations, and load the data into Java in a format that is conducive to conducting an analysis of the data. The simulation of the algorithms was carried out using the following hardware and software configurations, which are listed below for your reference: Laptop produced by Hewlett Packard (HP) with an Intel Duo Core processor running at 2.40GHz, 4.00GB of RAM, and a hard disc capacity of 500GB Software, the Windows 8.1 operating system, the Eclipse 8.0.1 integrated development environment (IDE), the Java Development Kit (JDK) 7.0.400.43, and the Pentaho Data Integration Tool were also included.

A. Treatment of Data Integrity for the Analysis

The process of handling the data, the classification used for the tax payer and the segmentation within the dataset is the key to ensuring the integrity of the analysis. This has made it possible to create risk profiles and differentiate between taxpayers who pose a high risk and those who pose a low risk. This results in better decisions being made regarding the targeting of the riskiest cases while placing less of a burden on businesses that are largely compliant and pose less risk. Then, decisions and strategies can centre on a graduated response to compliance behaviour, which makes compliance easier for those who want to comply while applying credible enforcement to those who do not comply. The classification of the organisation taxpayer dataset using SMO is displayed in Figures4.6. It consists of performance metrics like Kappa statistics and classified instances, for example. A build time of 223 milliseconds was accomplished, and 98.28 percent of the taxpayers were correctly classified.

The organisation taxpayer dataset that was classified using PSMO and the results that were obtained after classification indicated that 98.35 percent of taxpayers were correctly classified, and a build time of 223 milliseconds was achieved. While the classification of organisation taxpayer dataset that used Combined SMO showed that a Kappa statistics rate of 0.916 was obtained, which is 8% higher than the SMO and 7.8% higher than the PSMO, respectively. A build time of 223 milliseconds was accomplished, and 99.74 percent of the taxpayers were correctly classified. In the category of tax payers known as Individual Direct Assessment, correlative analyses of age, gender, source of income, income level, and number of years are performed against compliance. Additionally, the correlations of staff strength, years in business, business type, gross income, and income level are taken into consideration when assessing compliance for organisations that pay taxes using the Pay As You Earn (PAYE) system. When run through the model, some of the relationships between attributes and compliance are illustrated in the following figures.

According to the individual taxpayers, those with higher incomes are less likely to comply with the law in comparison to their counterparts in the middle and lower income categories. Taxpayers who have an untraceable source of income (also known as unknown income), such as cash, tips, or income from another source, are likely to be in violation of the law. It has also been observed that public limited liability companies have a higher compliance rate than the other business types, followed by private companies with a higher compliance rate than public limited liability companies. This can be explained by the fact that their taxes are deducted and paid automatically on a predetermined schedule each pay period. The passage of time in business has demonstrated that companies that have been operational for more than five years tend to be more compliant with regulations than businesses that are still relatively new and have been operating for fewer than five years.

IV. RESULTS AND DISCUSSION

The combined PSMO model that we have proposed can determine with a high level of accuracy whether a tax payer is likely to be compliant or not. It is able to make a prediction regarding the likelihood of defaulters taking into consideration the data that is input into the model. The receiver operating characteristic, also known as ROC, is a graphical representation of the ratio between the proportion of accurate positive diagnoses (also known as the true positive rate, or TPR) and the proportion of inaccurate negative diagnoses (also known as the false positive rate, or FPR). Tools such as Kappa Statistics and ROC curves and Confusion Matrix are utilised during the measurement process. The True Positive rate, the False Positive rate, the Precision, and the recall were some of the other metrics of measurement that were utilised. We made use of kappa statistics and confusion matrices because these tools measure the degree of agreement across major categories. When it comes to precision, it has been demonstrated that CSMO is superior to both SMO and PSMO.

The CSMO model's confusion matrix, which includes 8397 tax payers who have been correctly classified as non-compliant and 64 non-compliant There were taxpayers who were incorrectly classified as compliant, 958 taxpayers who were classified correctly as compliant, and there were no compliant taxpayers who were incorrectly classified as non-compliant. Kappa statistics is a measure of agreement between classifications and true classes that takes into account the influence of chance. The calculation for it involves subtracting the degree of agreement that would be expected to occur by chance from the degree of agreement that actually occurred, and then dividing that result by the greatest degree of agreement that could occur. The Kappa Statistics measure for the CSMO was 0.916, which is 8% higher than the measure for the SMO and 7.8% higher than the measure for the PSMO.

The correctness of the classification of the instances was the criterion that was used. When compared to the classification rates of SMO (98.28 percent) and parallel SMO (98.35 percent), CSMO's rate of correctly classified instances was significantly higher, coming in at 99.74 percent. The percentage of incorrectly classified instances of CSMO was recorded at 0.25 percent, which is a significantly lower value than the 1.72 percent of SMO and the 1.68 percent of PSMO. The Receiver Operating Characteristics Curve was obtained for SMO, PSMO, and CSMO by plotting the fraction of true positives out of the positives (TPR = true positive rate) versus the fraction of false positives out of the negatives (FPR = false positive rate). This allowed for the comparison of the true positive rate to the false positive rate. When compared to SMO and PSMO, the value that was achieved in CSMO, which was 0.9981, was superior.

Table 2 shows the performance of classifiers generated from the model. CSMO is compared with SMO and parallel SMO algorithms.

Table 2
Predictive Performance of Classifiers

Classifiers / Metrics	SMO	PSMO	CSMO
Correctly Classified Instances	98.28%	98.35%	99.74%
Incorrectly Classified Instances	1.72%	1.68%	0.25%
Time Taken to build Model	286ms	378ms	223ms
Relative Absolute Error	9.90%	10.33%	10.00%
Root Mean Square Value	0.13	0.13	0.13
Kappa Statistics	0.9124	0.916	0.994
Area below ROC Curve	0.944	0.913	0.9981
Confusion Matrix A= 0 B= 1 Total instances: 9419	Predicted ` Class: A, B Actual Class A: 8244, 217 Actual Class B: 0, 958	Predicted Class: A, B Actual Class A: 8293, 168 Actual Class B: 0, 958	Predicted Class: A, B Actual Class A: 8397, 64 Actual Class B: 0, 958

According to the findings, our algorithm performs better than the studies that were listed in terms of the accuracy of classification and prediction, the cost of data compilation, the avoidance of outliers, and validation. It does this by employing a more effective methodology in the building of a model, which will result in increased tax compliance. It allows seamless analytics without offloading from the database to the server, and it runs datasets in parallel processors for higher speed up, all of which contribute to it being a more cost-effective solution for the compilation of data than the study conducted by Butler [18]. Our innovative method demonstrates that it is better able to handle the drawback of over-fitting in Neural Networks in the Gonzalez & Velasquez [19] work because it strategically selects small data points to train on. This was demonstrated by the fact that they were able to train on. The cross-validation technique has not been implemented, and the test set has relied on the recommendations instead. According to research that was presented in the study even though Sequential Minimal Optimisation is considered to be the most effective, it requires a significant amount of computation time for the resolution of problems with a large size. As a result, our model reduces the impact of this disadvantage.

It emanates the same concept of combination that was used in the study that was conducted by Ferdousy et al. [10], but instead of sequentially, it makes an improvement by using two algorithms (SMO and PSMO) on two parts of the real dataset separately and then combining them to produce a robust model. This concept of combination was used in the study that was conducted by Ferdousy et al. [10]. The research provides a novel strategy for Eliminating the Tax Gap, Increasing the Level of Tax Compliance, and Detecting Patterns of Fraud. The turnaround time for processing returns and updating taxpayer records can also be cut down with this method.

Our system has presented visual data showing patterns of compliance, identified tax-payers as high risk or low risk, made predictions on tax-payer payments, and identified fraudulent behaviour on the part of tax-payers. This will help in reducing the tax gap, increase revenue, and provide a framework for decision support, all of which will assist BIRS in achieving its ultimate goal of increasing internally generated revenue (IGR). It is strongly suggested that the cost of splitting should be kept to a minimum during the initialization phase of the CSMO algorithm. It is recommended that additional methods of visual clustering analysis be incorporated into the system. In order to improve the root mean square value and relative absolute error rates of the CSMO model, it is recommended that a scenario involving a larger dataset be experimented on the model. This will allow for more iterations to be tested. Given that split (, 0) indicates only SMO, split (, 1) indicates only PSMO, and split (, 0.7) indicates only CSMO (Splitting dataset into two and running SMO on 40 percent and PSMO on 70 percent of the dataset).

In order to obtain varying degrees of output from the experiments, it is recommended to parameterize the split and conduct experiments using the split value of (, 0.5).

V. CONCLUSION

This paper has presented a function-based prediction technique that, when combined with the Sequential Minimal Optimisation Technique and the Parallel Sequential Minimal Optimisation Technique, significantly improves the accuracy of both classification and prediction. The correlation, classification, and prediction capabilities of the 775 Organization and 9419 Individual Taxpayer datasets were investigated. The metrics that were utilised in the analysis of the model revealed that the classification and prediction accuracy of the model is superior to that of the other works in the literature that were taken into consideration. The Kappa Statistics measure for the CSMO was 0.916, which is 8% higher than the measure for the SMO and 7.8% higher than the measure for the PSMO. Comparatively, our model had a classification accuracy of 99.74 percent when compared to the SMO's 98.28 percent and the parallel SMO's 98.35 percent. The percentage of incorrectly classified instances of CSMO was recorded at 0.25 percent, which is a significantly lower value than the 1.72 percent of SMO and the 1.68 percent of PSMO. The time it takes to build the model is 223 milliseconds, whereas in SMO it takes 286 milliseconds and in PSMO it takes 378 milliseconds. In the ROC Curve plot, the value obtained for CSMO was 0.9981, which was superior to the values obtained for SMO (0.944) and PSMO (0.913).

ACKNOWLEDGEMENT

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing

CONFLICT OF INTEREST

The author(s) declare that there is no conflict of Interest

REFERENCES

[1] A. D. Zubairu. Understanding Taxation. Nassarawa: Press, 2011..

[2] J. Abiola. & M. Asiweh. Impact of tax administration on government revenue in a developing economy-a case study of Nigeria. International Journal of business and social science, 3(8), 2012.

[3] S. Taddesse. Taxation for Sustained Prosperity: Collaborative Tax Policy Setting & Administration. Bublish, 2022, Inc.

[4] D. Desiderio. Data analysis techniques for enhancing the performance of Customs. World Customs Journal, 13(2), 2019, pp.17-22.

[5] M. Önder, & U.N. Zengin. A framework for comparative analysis: Public administration across the globe. In The Palgrave Handbook of Comparative Public Administration (pp. 51-76). 2022, Palgrave Macmillan, Singapore.

[6] T. Connolly, & C. Begg. Database Systems. A Practical Approach to Design, Implementation and Management. Harlow: 2015, Addisson Wesley.

[7] P. Simon. Too Big to Ignore: The Business Case for Big Data. 2013, Wiley.

[8] T.M. Mitchell. Machine Learning. 1995, McGraw-Hill

[9] C.S. Lee, Cheang, P.Y.S. and M. Moslehpour. Predictive analytics in business analytics: decision tree. Advances in Decision Sciences, 26(1), 2022, pp.1-29.

[10] E.Z. Ferdousy, M.M. Islam & M. Abdul Matin. Combination of Naive Bayes Classifier and K-Nearest Neighbour (cNK). Computer and Information Science, 6, 2013, 48-56.

[11] M. Kachele, G. Palm, G., & F. Schwenker. SMO Lattices for the Parallel Training of Support Vector Machines. European Symposium on Artificial Neural Network (pp. 579-584), 2015, Ulm: Bruges.

[12] S. Singh. Digital Transformation in the Tax World. Tax Executive, 2022, 74, p.64.

[13] Junquera-Varela, R.F., Lucas, C.O., Krsul, I., Calderon Yksic, V.O. and Arce Rodriguez, P., 2022. Digital Transformation of Tax and Customs Administrations.

[14] J. Owens, B. Schlenther, K.S. Adjeyi, K.E. Amaglo, T. Kouevi, A. Jansen, O.K. Mackenzie. Steyn, W., Tchila, P., Okey, M.K.N. and N. Traore. Developments in the Use of Technologies in African Tax Administrations. African Multidisciplinary Tax Journal, 2022(1), pp.1-22.

[15] N. Noor, A. Sarlan, and N. Aziz. Revenue Prediction for Malaysian Federal Government Using Machine Learning Technique. In 2022 11th International Conference on Software and Computer Applications, 2022, pp. 143-148.

[16] J. Choe, A. Crotty. T. Moreshet. M. Herlihy,. and R.I. Bahar.. HybriDS: Cache-Conscious Concurrent Data Structures for Near-Memory Processing Architectures. In Proceedings of the 34th ACM Symposium on Parallelism in Algorithms and Architectures, 2022, pp. 321-332.

[17] Z. Chu, P. Xiao, P., Mi, D., Hao, W., Khalily, M. and L.L. Yang.. A novel transmission policy for intelligent reflecting surface assisted wireless powered sensor networks. IEEE Journal of Selected Topics in Signal Processing, 15(5), 2021, pp.1143-1158.

[18] J. Butler. Big Data and Analytics at the IRS. Proceedings of IBM-TechAmerica , 6, 2012.

[19] P.C. Gonzalez, & J.D. Velasquez. Characterization and detection of tax payers with false invoices using data mining techniques. Expert Systems with Applications , 1. 2013.