

# A Comparison of SOM and K-Means Algorithms in Predicting Tax Compliance

**A****Felix Bankole***University of South Africa, South Africa***Zama Vara***University of South Africa, South Africa*

## INTRODUCTION

Revenue administrations oversee the collection and management of domestic revenues such as taxes and customs duties. Tax compliance is a major problem for tax administrations across the world (Jenkins, 2018). Tax noncompliance, is defined as a failure to file on time, capturing incomplete or misleading information on returns, and overdue payment of taxes (Jenkins, 2018). Furthermore, tax noncompliance can be explained as a fraudulent or deliberate misrepresentation of information on income tax returns to decrease the amount of the tax liability (Vanhoeuyveld, Martens, & Peeters, 2019). Tax noncompliance is also characterised by false claims for refund, misclassification of goods and services, bogus traders, under-reported sales, and failure to register and failure to pay taxes to the tax administration (Bimo, Prasetyo, & Susilandari, 2019).

The payment of taxes is of significant concern to lawmakers in developed and developing nations. Tax administration in countries worldwide have utilized different methods to quantify the nature of non-compliance, which depends on the social and economic behavior of citizens (Agyeiwa, et al., 2019). In the opinion of Krishnamurthya & Desouza (2014), fiscal revenue deficiencies due to tax noncompliance are very considerable in third world countries of the sub-Saharan Africa, Latin America, Caribbean, and South Asia regions. Sadly, these are the countries which have a greater reliance on tax revenues for their fiscal planning (Krishnamurthya & Desouza, 2014). Compounding the problem is that these countries are the most affected by budget shortages. Moreover, tax administrations are responsible for investing in efficient and effective techniques to select taxpayers for audit. Tax compliance and fraud detection is likely to become even more important with recent developments in Artificial Intelligence (Krishnamurthya & Desouza, 2014).

Auditing tax returns is a slow and costly process that is very prone to errors. Conducting tax audits for example, involves costs to the tax administration, as well as to the taxpayer (Zwick, 2021). Furthermore, the field of anomaly and fraud detection is characterised by unlabelled historical data (Thudumu, Branch, Jin, & Singh, 2020). To this end, here the use of unsupervised machine learning algorithms is suggested. They are well suited to unlabelled datasets. Notwithstanding there is little research on comparing the effectiveness of various unsupervised learning approaches in the income tax realm. In this paper, two clustering techniques are proposed. These are, a K-means algorithm and Self-Organizing Map. Moreover, the efficacy and the predictive accuracy of a Self-Organizing Map (SOM) and K-means algorithms in identifying income tax noncompliance are compared.

DOI: 10.4018/978-1-7998-9220-5.ch155

In this study a SOM and K-means are used to obtain taxpayer cohorts that can easily be labelled as either compliant or otherwise. The objective of this research is to firstly create an Artificial Intelligence based, tax compliance detection framework and secondly, to compare a Self-Organizing Map (SOM) with a K-means algorithm to detect income tax noncompliance. To delineate, this study will utilize taxpayer datasets of the telecommunications industry. Furthermore, to confine the study, unsupervised learning approaches are chosen on income tax noncompliance. Additionally, this study will only use taxpayer datasets of the telecommunications industry.

## **BACKGROUND**

Historical research on tax compliance has been met with a lack of historical data (Battiston & Gamba, 2016). Tax compliance cannot be attributed entirely by the level of enforcement, rather it is also based on behavioural factors affecting the purchaser-seller relations (Lamantia & Pezzino, 2021). Researchers have made attempts to identify and measure the impact of tax morale on tax compliance. Luttmer and Singhal (2014) describe various mechanisms according to which tax morale may affect compliance. One mechanism is based on intrinsic motivation. In every society, there are individuals who believe that paying taxes is a way to contribute to society's welfare and that individuals may obtain private utility from it. There is another important aspect of taxpayers' behaviour related to tax morale and that is the role those social norms and reputation may play (Luttmer & Singhal, 2014).

Tax compliance experiments have been gaining impetus in recent years. The recent phenomenon of an increased number of studies on tax compliance is due to an increase in data availability. These contemporary developments have allowed significant advances in the investigational literature on tax compliance (Mascagni, 2018). As a consequence, tax administrations are investing significant resources in the development of artificial intelligence algorithms to improve tax compliance, according to a new OECD report (OECD, 2021). John McCarthy first coined the term "Artificial Intelligence". He defines it as the science and engineering of making intelligent machines (Rajaraman, 2016). In their study, Faúndez-Ugaldea et al. (2020), consider two applications of artificial intelligence, that is, taxpayers' risk and the automation of tax audit case selection. However, extraordinarily little has been detailed in the literature on how AI, exist side-by-side with taxpayers' rights. Faúndez-Ugaldea et al. (2020) report that in some countries the access to these algorithms is not clearly regulated. Furthermore, they note that general principles derived from the fundamental rights proclaimed by each country make it possible to safeguard taxpayers' right to access this information (Faúndez-Ugaldea et al, 2020).

## **FOCUS OF THE ARTICLE**

### **Tax Compliance**

Tax compliance refers to a taxpayer's decision to comply with tax laws and regulations by paying taxes when expected, filing returns timely and accurately (Youde & S.Lim, 2019). Factors that affecting voluntary tax compliance can be divided into five major groupings, that is, demographic factors like age, gender and education; individual factors like tax knowledge, personal financial constraints and awareness of offense and penalty (Saxunova & Szarkova, 2018); social attributes such as the perception on equity and fairness of tax system; institutional factors such as the simplicity of taxation, role of tax

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-comparison-of-som-and-k-means-algorithms-in-predicting-tax-compliance/317698](http://www.igi-global.com/chapter/a-comparison-of-som-and-k-means-algorithms-in-predicting-tax-compliance/317698)

## Related Content

---

### Development, Monitoring, and Management Approaches of Machine Learning Implementations for the Effective Delivery of Government Services

Santosh Ramkrishna Durugkar (2024). *Machine Learning and Data Science Techniques for Effective Government Service Delivery* (pp. 225-253).

[www.irma-international.org/chapter/development-monitoring-and-management-approaches-of-machine-learning-implementations-for-the-effective-delivery-of-government-services/343116](http://www.irma-international.org/chapter/development-monitoring-and-management-approaches-of-machine-learning-implementations-for-the-effective-delivery-of-government-services/343116)

### Machine Learning for Web Proxy Analytics

Mark Maldonado and Ayad Barsoum (2022). *Research Anthology on Machine Learning Techniques, Methods, and Applications* (pp. 870-881).

[www.irma-international.org/chapter/machine-learning-for-web-proxy-analytics/307488](http://www.irma-international.org/chapter/machine-learning-for-web-proxy-analytics/307488)

### Foundational Recommender Systems for Business

Prageet Aeron (2023). *Encyclopedia of Data Science and Machine Learning* (pp. 2799-2816).

[www.irma-international.org/chapter/foundational-recommender-systems-for-business/317713](http://www.irma-international.org/chapter/foundational-recommender-systems-for-business/317713)

### Machine Learning in Authentication of Digital Audio Recordings

Rashmika Kiran Patole and Priti Paresh Rege (2020). *Machine Learning and Deep Learning in Real-Time Applications* (pp. 145-167).

[www.irma-international.org/chapter/machine-learning-in-authentication-of-digital-audio-recordings/257318](http://www.irma-international.org/chapter/machine-learning-in-authentication-of-digital-audio-recordings/257318)

### Forecasting Price of Amazon Spot Instances Using Machine Learning

Manas Malik and Nirbhay Bagmar (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 71-82).

[www.irma-international.org/article/forecasting-price-of-amazon-spot-instances-using-machine-learning/277435](http://www.irma-international.org/article/forecasting-price-of-amazon-spot-instances-using-machine-learning/277435)