Abstract

Machine learning (ML) is an application of artificial intelligence (AI). It provides systems the ability to learn and improve automatically from experiences. It uses supervised, unsupervised, and reinforcement strategies to learn multi-level representations for the tasks of classification.

Tax fraud is a general term that refers to the efforts of organizations or individuals to legally defraud, such as concealing the true status of the taxpayer to the tax authorities so that the value of the tax is reduced and includes, for example, the submission of false tax reports, such as declaring earnings that are undervalued. In other words, tax fraud is lying on a tax return form to reduce tax liability. Therefore, detecting tax fraud is one of the main priorities of the tax authorities. Most of the recent works and modern businesses in tax fraud detection in many countries around the world rely on machine learning techniques that make use of labeled or audit-assisted data.

In this thesis, the researcher aimed to build a new ML model that is used for detecting Tax levels of tax evasion based on a hybrid neural network. The proposed model guarantees the accuracy and efficiency of the results. The proposed model is applied in a real case application for detecting tax fraud in the Tax Authority of Yemen to reduce the effort, time, and cost spent in trying to reduce tax evasion and fraud. The datasets used for validation and verification of the proposed model are given by the Tax Authority in Yemen.

Keywords: Machine learning, Hybrid Learning, Tax fraud, and Taxpayer.

المستخلص

التعلم الآلي هو تطبيق للذكاء الاصطناعي. فهو يوفر للأنظمة القدرة على التعلم والتطور تلقائيًا من التجارب. ويستخدم استراتيجيات التعلم الخاضع للإشراف والغير خاضع للإشراف والتعزيز لتعلم التمثيلات متعددة المستويات لمهام التصنيف.

الاحتيال الضريبي هو مصطلح عام يشير إلى الجهود التي تبذلها المنظمات أو الأفراد للاحتيال قانونيا، مثل إخفاء الوضع الحقيقي لدافع الضرائب أمام السلطات الضريبية بحيث يتم تخفيض قيمة الضريبة ويشمل ذلك، على سبيل المثال، تقديم تقارير ضريبية كاذبة، مثل الإعلان عن الأرباح المقومة بأقل من قيمتها الحقيقية. بمعنى آخر، الاحتيال الضريبي هو الكذب على نموذج الإقرار الضريبي لتقليل الالتزام الضريبي. ولذلك، يعد اكتشاف الاحتيال الضريبي أحد الأولويات الرئيسية للسلطات الضريبية. تعتمد معظم الأعمال الحديثة في مجال كشف الاحتيال الضريبي في العديد من البلدان حول العالم على تقنيات التعلم الآلي التي تستفيد من البيانات المصنفة أو البيانات المدعومة بالتدقيق.

تهدف الباحثة في هذه الأطروحة إلى بناء نموذج جديد للتعلم الآلي يستخدم للكشف عن مستويات التهرب الضريبي بالاعتماد على الشبكة العصبية الهجينة. ويضمن النموذج المقترح الدقة والكفاءة في النتائج. تم تطبيق النموذج المقترح على بيانات حقيقية مقدمة من مصلحة الضرائب اليمنية لتقليل الجهد والوقت والتكلفة المبذولة في محاولة الحد من التهرب والاحتيال الضريبي. مجموعات البيانات المستخدمة للتحقق من النموذج المقترح مقدمة من مصلحة الضرائب اليمنية.

الكلمات المفتاحية: التعلم الآلي، التعلم الهجين، الاحتيال الضريبي، ودافعي الضرائب.