MODELO PARA DETECTAR LA EVASIÓN DEL IMPUESTO A LA RENTA EN EL SECTOR DE LOS PRINCIPALES CONTRIBUYENTES PERUANOS.

por

Frank Jonislla Pillaca

Avance de informe de tesis en cumplimiento parcial

de los requisitos para el grado

Magister en Ingeniería de Sistemas e Informática

mención Ingeniería de software.

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

Diciembre 2018

TABLE OF CONTENTS

[LIST OF TABLES v](#_Toc532988864)

[LIST OF FIGURES vi](#_Toc532988865)

[CAPITULO 1: INTRODUCCIÓN 7](#_Toc532988866)

[**Antecedentes** 7](#_Toc532988867)

[**Problema** 9](#_Toc532988868)

[**Importancia del problema** 9](#_Toc532988869)

[**Motivación** 12](#_Toc532988870)

[**Objetivos** 12](#_Toc532988871)

[CHAPTER 2: REVISIÓN DE ARTICULOS 13](#_Toc532988872)

[**Metodología de Investigación** 13](#_Toc532988873)

[Planificación de la revisión 13](#_Toc532988874)

[Desarrollo de la revisión. 14](#_Toc532988875)

[Resultado de la revisión. 16](#_Toc532988876)

[**Análisis** 19](#_Toc532988877)

[¿Qué modelos existen para predecir la evasión de impuestos? 19](#_Toc532988878)

[¿Qué algoritmos de machine learning existen para predecir la evasión de impuesto? 21](#_Toc532988879)

[¿Qué métodos se usan para medir la efectividad de un modelo de evasión de impuestos? 22](#_Toc532988880)

[REFERENCIAS 24](#_Toc532988881)

##### LIST OF TABLES

[Tabla 1 *Criterios de inclusión de los artículos seleccionados* 14](#_Toc532988882)

[Tabla 2 *Lista de criterios de exclusión de artículos.* 14](#_Toc532988883)

[Tabla 3 *Modelos de evasión de impuestos.* 19](#_Toc532988884)

[Tabla 4 *Algoritmos usados para detectar la evasión de impuestos.* 21](#_Toc532988885)

[Tabla 5 *Algoritmos usados para detectar la evasión de impuestos.* 22](#_Toc532988886)

##### LIST OF FIGURES

[Figura 1. *Estimación de la evasión tributaria de los clientes informales del sistema financiero Fuente(http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2016/documento-de-trabajo-11-2016.pdf)* 10](#_Toc532988887)

[Figura 2. *Proceso de selección de artículos.* 16](#_Toc532988888)

[Figura 3. *Tendencia de los artículos seleccionados por años.* 17](#_Toc532988889)

[Figura 4. *Tendencia de los artículos seleccionados por años.* 17](#_Toc532988890)

[Figura 5. *Distribución de artículos por países.* 18](#_Toc532988891)

[Figura 6. *Distribución de artículos por área de conocimiento.* 18](#_Toc532988892)

[Figura 7. *Clasificación de algoritmos de aprendizaje según el tipo de aprendizaje* (Qiu et al. 2016)*.* 21](#_Toc532988893)

##### CAPITULO 1: INTRODUCCIÓN

# **Antecedentes**

**¿Qué son los impuestos?**

“Existen necesidades comunes a mucha gente cuya satisfacción exige esfuerzo colectivo, porque cada persona por separado no tendría capacidad para financiarlas”. Para satisfacer estas necesidades colectivas, existen administraciones públicas, como la SUNAT, que se encargan de planificar los gastos de los servicios públicos que disfrutan los ciudadanos. Pero para poder pagar estos gastos, las administraciones públicas tienen que obtener unos ingresos que proceden de los ciudadanos; estos ingresos son los IMPUESTOS”.

**Características de los impuestos**

Pueden ser: Permanentes o transitorios; Directos o indirectos, proporcionales, progresivos, regresivos. Por ejemplo, el impuesto de renta es considerado un impuesto progresivo en la medida en que los agentes con mayores ingresos deben pagar una tasa más alta. Por el contrario, un impuesto regresivo es aquel en el que se capta un porcentaje menor en la medida en que el ingreso aumenta.

Pueden ser reales o personales; Impuestos reales son aquellos cuya naturaleza se determina con independencia del elemento personal de la relación tributaria, y pueden ser definidos por la norma sin referencia a ningún sujeto determinado y gravan una manifestación de riqueza sin ponerse en relación con una determinada persona.

Si todos como sociedad nos involucramos más activamente en los impuestos, seremos una sociedad más participativa en la generación de riqueza, en la aplicación de los recursos gubernamentales, en la exigencia de la transparencia y en la rendición de cuentas a nuestros servidores públicos.

Ejemplos de impuestos:

• Impuesto General a las Ventas

• Impuesto Extraordinario de Solidaridad

• Impuesto general a las rentas

• Impuesto a la Renta

**Impuesto a la Renta.**

El impuesto a la renta es un tributo que se determina anualmente y su ejercicio inicia el 01 de enero y finaliza el 31 de diciembre.

Este impuesto se aplica a los ingresos que provienen del arrendamiento u otro tipo de cesión de bienes muebles o inmuebles, acciones u otros valores mobiliarios, y/o del trabajo realizado de forma dependiente o independiente.

Para la determinación del Impuesto a la Renta de Personas Naturales que no realizan actividad empresarial, corresponde lo siguiente:

Rentas de Capital - Primera Categoría: generadas por el arrendamiento, subarrendamiento o cualquier tipo de cesión de bienes muebles o inmuebles. El pago debes hacerlo tú mismo y corresponde al 6.25% sobre el monto obtenido por la renta.

Rentas de Capital - Segunda Categoría: corresponde a los intereses por colocación de capitales, regalías, patentes, rentas vitalicias, derechos de llave y otros.

Tercera categoría: las derivadas de actividades comerciales, industriales, servicios o negocios.

Rentas del Trabajo (Cuarta y/o Quinta categoría) y Renta de Fuente Extranjera: Si eres trabajador independiente o dependiente y tu ingreso anual es superior a 7UIT S/. deberás pagar impuesto a la renta.

En el Perú el impuesto a la renta es el que más se evade.

Evasión Fiscal.

Existe evasión fiscal cuando una persona infringiendo la ley, deja de pagar todo o una parte de un impuesto al que está obligada. Al incumplir de manera intencional con el pago de las contribuciones que le corresponden como contribuyente y ciudadano, estará cometiendo un delito. Este incumplimiento causa un deterioro económico en las Finanzas Públicas y el contribuyente obtiene un beneficio que no le correspondería, razón por la cual incurre en el delito de defraudación fiscal; para detener este flagelo es necesario generar herramientas que permitan detectar la evasión tributaria, principalmente la evasión del impuesto a la renta.

# **Problema**

La detección de empresas, del sector de los principales contribuyentes peruanos, que no registran sus ganancias de forma completa y por ende evaden el impuesto a la renta.

# **Importancia del problema**

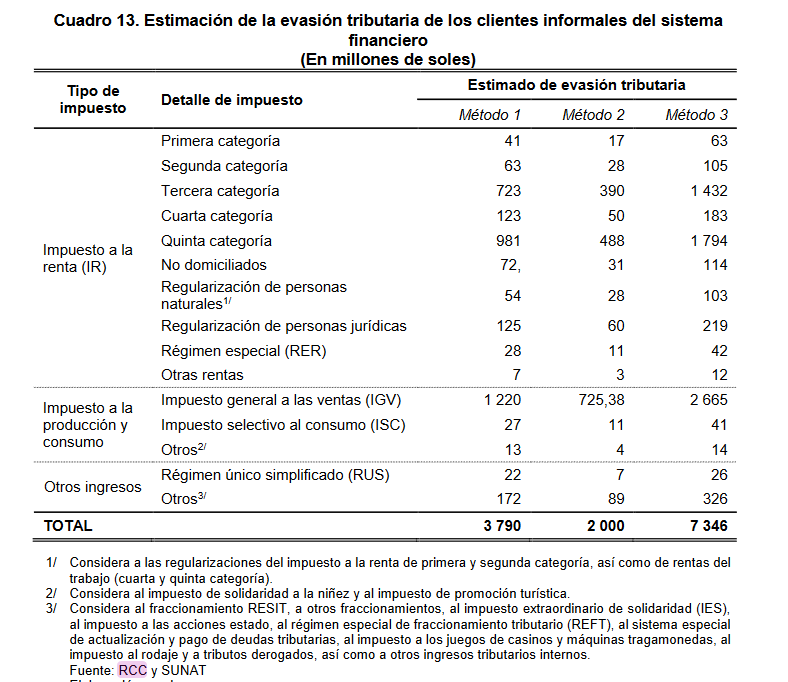
El fraude tributario es un fenómeno mundial que afecta a todas las economías del planeta en mayor o menor medida con pérdidas económicas que van desde el 8 hasta el 63% del PBI dependiendo de cada país analizado (Banco Mundial, 2017). En el Perú contamos con una economía altamente informal donde 8 de cada 10 dólares se mueven fuera del sistema financiero y operan principalmente en mercado con el uso de efectivo que se inserta al sistema y evade impuestos, no es rastreable y no opera de acuerdo con el contexto formal de la economía. Existen cerca de 6.5 millones de empresas y 10 millones de empresarios individuales, pero solo 12 700 empresas representan el 75% del total de la recaudación tributaria nacional.

La evasión tributaria de IGV en el Perú alcanza el 36% del potencial de recaudación y la evasión del impuesto a la renta (IR) es del orden del 57% perdiéndose en general por el total de evasión tributaria cerca de 22 000 millones de dólares anuales que el Estado peruano deja de percibir por este problema (SUNAT, 2018).

En la figura 1 muestra los resultados de la estimación de la evasión tributaria de los clientes informales del sistema financiero, a nivel total y por tributo.

Tomando en cuenta la información de los tres métodos empleados, se observa que la evasión total podría ubicarse dentro del rango de S/ 2 000 millones (0,3 por ciento del PBI) y S/ 7 346 millones (1,3 por ciento del PBI), aproximadamente.

Figura 1. *Estimación de la evasión tributaria de los clientes informales del sistema financiero Fuente(*[*http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2016/documento-de-trabajo-11-2016.pdf*](http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2016/documento-de-trabajo-11-2016.pdf)*)*



# **Motivación**

La tarea de detección de fraude no es un tema fácil de resolver, teniendo en cuenta las múltiples modalidades y evolución rápida que este tema ha tenido en la actualidad, las entidades de administración tributaria a nivel mundial utilizan la ciencia de la estadística con herramientas de la minería de datos y el machine learning para reconocimiento de patrones de comportamiento fraudulento. Sin embargo, a pesar del progreso sustancial, en esta área, los métodos disponibles presentan deficiencias que limitan su utilidad, a veces debido a hipótesis poco claras que subyacen al método. Lo más probable es que esto continúe para siempre, ya que es bien sabido que la imaginación de los delincuentes conduce a una manipulación más sofisticada.

# **Objetivos**

**Objetivo general**: Diseñar un modelo de detección de evasión de impuestos usando técnicas de machine learning.

**Objetivo específico**: Diseñar un modelo que sea capaz de seleccionar contribuyentes que no estén cumpliendo sus deberes tributarios

##### CHAPTER 2: REVISIÓN DE ARTICULOS

# **Metodología de Investigación**

Este estudio a seguido una secuencia de pasos sistemático y simple, es capaz de ser reproducido sin problemas y de manera intuitiva, esta metodología consta de 3 etapas.

Planificación de la revisión: En esta fase, se plantean las preguntas de investigación y se define el protocolo de revisión.

Realización de la revisión: en esta fase, el plan se ejecuta y se seleccionan los artículos primarios, de acuerdo con los criterios de inclusión y exclusión establecidos para el efecto.

Resultado de la revisión: En esta fase, se muestran las estadísticas y el análisis de los documentos encontrados y seleccionados, y que se analizan a continuación.

### Planificación de la revisión

Para responder a la pregunta de investigación, se hacen las siguientes preguntas sobre las técnicas y herramientas necesarias para detectar la evasión de impuestos.

P1: ¿Qué métodos existen para predecir la evasión de impuestos?

P2: ¿Qué algoritmos de machine learning existen para predecir la evasión de impuesto?

P3: ¿Qué métodos se usan para medir la efectividad de un modelo de evasión de impuestos?

La fuente de búsqueda es proporcionada por el metabuscador: “Web of Science”, el cual es un servicio en línea de información científica. El período de búsqueda comienza en el año 2016, ya que es un periodo razonable para contar con información reciente y actualizada; puesto que la tecnología avanza de forma rápida.

Usamos la siguiente cadena de búsqueda en los títulos, resumen y palabras clave: ts=(((tax AND income) OR (tax AND evasion) OR (tax AND avoidance) OR (tax AND default) OR( tax AND compliance)) AND ((big AND data) OR (machine AND learning) OR (neural AND network) OR (genetic AND algorithm))).

### Desarrollo de la revisión.

Los criterios de inclusión y exclusión se han considerado, como se muestra en las Tablas 1 y 2, respectivamente.

Tabla 1 *Criterios de inclusión de los artículos seleccionados*

|  |  |
| --- | --- |
| **Criterios de inclusión** | |
| **Criterio de inclusión** | **Razón de inclusión** |
| Objetivo de la investigación | Estudios que permiten predecir la evasión de impuestos. |
| Estudios empíricos cuantitativos. | Estos artículos se incluyen porque proporcionan evidencia empírica existente, lo que representa el interés principal de esta revisión. |
| Lenguaje ingles | Sólo artículos en ingles fueron considerados. |

Tabla 2 *Lista de criterios de exclusión de artículos.*

|  |  |
| --- | --- |
| **Criterios de exclusión** | |
| **Criterio de exclusión** | **Razón de exclusión** |
| Tipo de publicación | Se excluyeron libros, capítulos de libros y disertaciones. |
| Unidad de Análisis | Se excluyeron estudios que no consideren soluciones basadas en machine learning. |
| Enfoque de la investigación | Se excluyeron estudios que no muestran metodología de investigación, pruebas numéricas (estadística descriptiva) y análisis o discusión. |

Como primer paso, las palabras clave y sus respectivas descripciones se utilizaron para buscar los artículos principales. La revisión se limitó a artículos en revistas revisadas por pares, omitiendo libros, capítulos de libros y actas de congresos, ya que los artículos de revistas se consideran un conocimiento válido y representan declaraciones autorizadas sobre el tema (Ardito et al., 2015). El procedimiento de búsqueda consideró las ediciones disponibles de revistas del período 2005. Se identificaron un total de 59 estudios potenciales, estos fueron sometidos a un proceso de selección de acuerdo con los criterios de inclusión y exclusión establecidos. Así, primero, dividimos los artículos en "artículos excluidos" (19 artículos) y "artículos para el chequeo de texto completo" (40 artículos), fue necesario realizar una revisión previa de los títulos y resúmenes. Sin embargo, muchos resúmenes no proporcionan una comprensión clara del propósito de los artículos, y se seleccionaron 20 referencias. A continuación, un análisis más detallado de los artículos de "verificación de texto completo", se procedió a leer la introducción y conclusiones, obteniendo 16 referencias. Finalmente, procedimos a leer el contenido completo del artículo para determinar su relevancia para el presente estudio y, principalmente, para determinar si estos estudios identifican los factores críticos de éxito, al notar que la cantidad de artículos no era suficiente se tuvo que recurrir a artículos presentados en conferencias que nos permitan tener una cantidad aceptable de documentos a tomar en cuenta para el estudio. Así, la muestra final consta de 16 estudios primarios y 4 de conferencias. Nos encontramos con un gran inconveniente al buscar artículos relacionados a la evasión de impuestos y aprendizaje automático.

Los procesos aplicados, así como los resultados obtenidos en cada paso del proceso, se representan en la Figura 2.

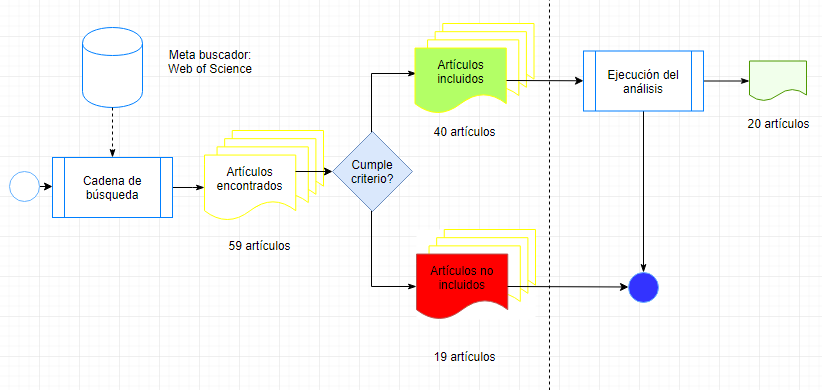


Figura 2. *Proceso de selección de artículos.*

### Resultado de la revisión.

El resultado del proceso de selección de los artículos nos dio 59, de los cuales 20 fueron seleccionados. Los cuales fueron analizados para responder las preguntas de investigación.

De acuerdo a las tendencias de citación de los artículos nos podemos dar cuenta que el área de evasión de impuestos y machine learning, no ha sido muy estudiado; la tendencia indica que a partir del 2017 los artículos referentes al estudio, vienen siendo referenciados.

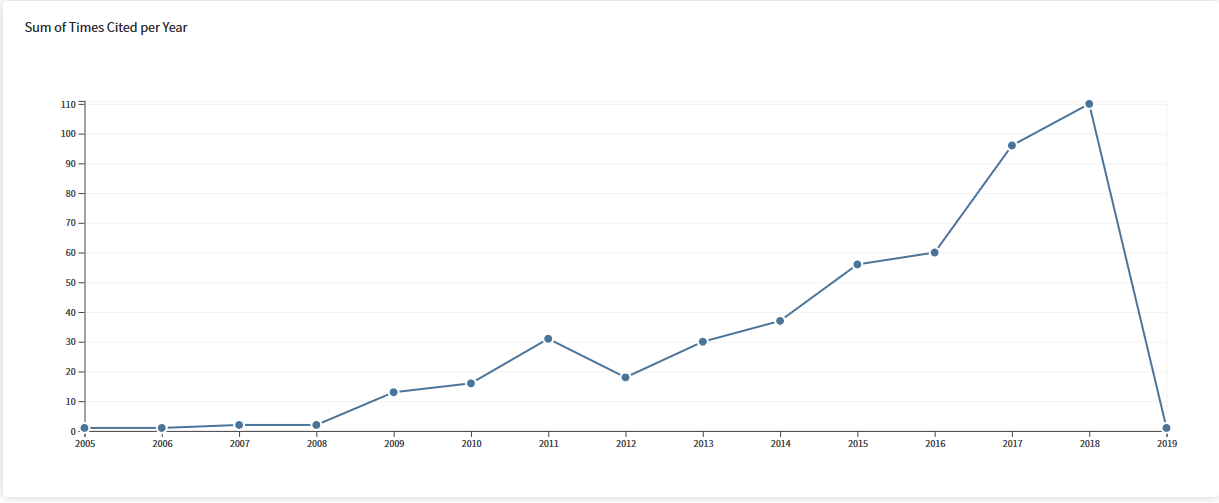


Figura 3. *Tendencia de los artículos seleccionados por años.*

La mayor cantidad de artículos, fueron encontrados en el año 2017, tal como en la figura 4, lo cual demuestra que el área de evasión tributaria y machine learning no muestra mucho interés en la academia.

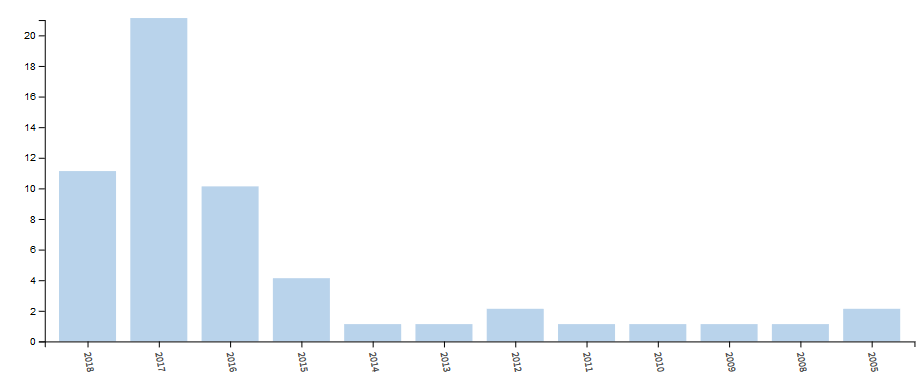


Figura 4. *Tendencia de los artículos seleccionados por años.*

Los países con mayor cantidad de artículos referentes al estudio son estados unidos e Inglaterra, puesto que los países con más ingresos son los que más perdidas tienen debido a la evasión de impuestos.



Figura 5. *Distribución de artículos por países.*

Como era de espera la mayor cantidad de artículos se encontraron en las áreas de economía, ingeniería y ciencias de la computación.

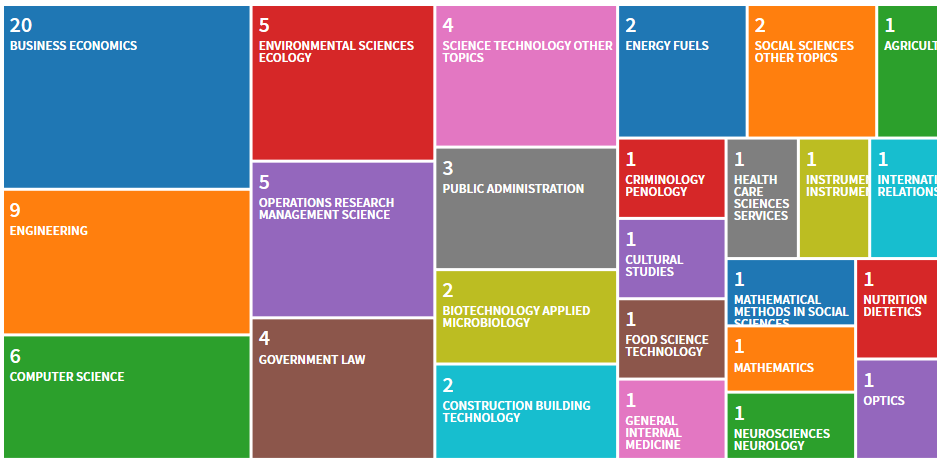


Figura 6. *Distribución de artículos por área de conocimiento.*

# **Análisis**

Esta sección responde las preguntas de investigación lanzadas en la etapa de planeamiento.

### ¿Qué modelos existen para predecir la evasión de impuestos?

En trabajos relacionados al modelo de detección de evasión de impuestos se pueden agrupar en dos categorías principales: i) analítico (macroeconómico y basado en el principio de agentes), y ii) computacional (basado en agentes, basado en la simulación) (Goumagias, Hristu-Varsakelis, and Assael 2018) en esta investigación nos centraremos en los modelos computacionales.

Tabla 3 *Modelos de evasión de impuestos.*

|  |  |
| --- | --- |
| **Modelo** | **Referencia** |
| modelo de co-evolución de los esquemas de Evasión tributarias con las políticas de auditoría. | (Hemberg et al. 2016) |
| Modelo de evaluación de informes financieros anulas, aplicando la ley de Benford. | (Ausloos, Cerqueti, and Mir 2017) |
| Dinámica de la evasión fiscal a través de un modelo de tipo epidémico; estudia la probabilidad de que un contribuyente cambie de comportamiento de acuerdo al comportamiento de sus relaciones. | (Brum and Crokidakis 2017) |
| Modelo para el estudio de la aversión al riesgo, para comprender la Evasión fiscal de las empresas. | (Goumagias, Hristu-Varsakelis, and Assael 2018) |
| Modelo para determinar un subconjunto óptimo o casi óptimo de variables para un modelo de análisis discriminante lineal. | (Höglund 2017) |
| Modelo para detectar Evasión tributaria usando un sistema inteligente hibrido. | (Rahimikia et al. 2017; Jupri and Sarno 2018; Mabe-Madisa 2018) |
| Modelo para predecir la Evasión de impuestos usando el proceso de regresión Gaussiana. | (Babu and Vasavi 2017) |
| Modelo de red de color para representar el comportamiento económico. | (Tian et al. 2016) |
| Modelo de análisis visual que da soporte a la detección de evasión de impuestos | (Didimo et al. 2018) |
| Modelo de predicción de evasión de impuestos mediante el análisis de redes sociales. | (Lismont et al. 2018) |
| Modelo de detección de fraude tributario usando técnicas de machine learning no supervisado. | (de Roux et al. 2018) |
| Modelo de detección de fraude inter-regiones usando transfer-learning. | (Zhu et al. 2018) |
| Modelo de detección de evasión de impuestos usando patrones de contraste. | (Zheng et al. 2016) |

### ¿Qué algoritmos de machine learning existen para predecir la evasión de impuesto?

El campo del aprendizaje automático se divide en tres subdominios: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo (Qiu et al. 2016).

El aprendizaje supervisado requiere capacitación con datos etiquetados que tienen entradas y salidas deseadas.

El aprendizaje no supervisado no requiere datos de capacitación etiquetados y el entorno solo proporciona entradas sin los objetivos deseados.

El aprendizaje por refuerzo permite aprender de los comentarios recibidos a través de interacciones con un entorno externo.

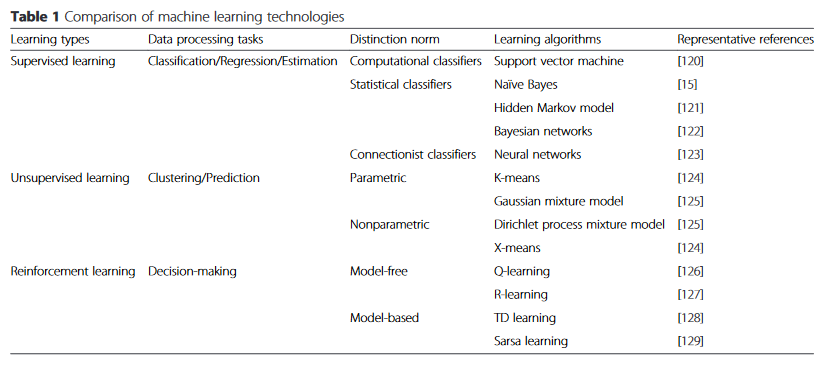


Figura 7. *Clasificación de algoritmos de aprendizaje según el tipo de aprendizaje* (Qiu et al. 2016)*.*

Tabla 4 *Algoritmos usados para detectar la evasión de impuestos.*

|  |  |
| --- | --- |
| **Algoritmo** | **Referencia** |
| MLP (Perceptrón multicapa) | (Jupri and Sarno 2018; Rahimikia et al. 2017) |
| Naive Bayes | (Jupri and Sarno 2018; Mabe-Madisa 2018) |
| SVM (Máquina de vectores de soporte) | (Jupri and Sarno 2018; Rahimikia et al. 2017) |
| KNN (Vecino más cercano K) | (Jupri and Sarno 2018) |
| C4.5 | (Jupri and Sarno 2018) |
| Transfer Component Analysis (TCA) | (Zhu et al. 2018) |
| LightGBM | (Zhu et al. 2018) |
| Transfer Adaboost | (Zhu et al. 2018) |
| árbol de decisión | (Lismont et al. 2018; Mabe-Madisa 2018) |
| Teoría de grafos | (Tian et al. 2016; Didimo et al. 2018) |
| La ley del primer dígito de Benford | (Ausloos, Cerqueti, and Mir 2017) |
| Redes neuronales profundas | (Goumagias, Hristu-Varsakelis, and Assael 2018) |
| Q-learning | (Goumagias, Hristu-Varsakelis, and Assael 2018) |
| Análisis discriminante lineal | (Höglund 2017) |
| Logistic regression (LR) | (Rahimikia et al. 2017) |
| Harmony search (HS) optimization algorithm | (Rahimikia et al. 2017) |
| Gaussian Regression process with varying hyper parameters | (Babu and Vasavi 2017) |
| Regresión logística | (Lismont et al. 2018) |
| Random forest | (Lismont et al. 2018) |

### ¿Qué métodos se usan para medir la efectividad de un modelo de evasión de impuestos?

Evaluar el desempeño, las debilidades y fortalezas de los modelos de predicción es muy importante. Para ello, se utiliza una métrica para seleccionar el modelo probabilístico que predice los mejores resultados en función de criterios como accuracy, precisión, recall, área bajo la curva ROC y F beta score.

Tabla 5 *Algoritmos usados para detectar la evasión de impuestos.*

|  |  |
| --- | --- |
| **Algoritmo** | **Referencia** |
| Exactitud | (Zheng et al. 2016; Qiu et al. 2016; Qiu et al. 2016; Lismont et al. 2018; Mabe-Madisa 2018; Jupri and Sarno 2018; Piolatto and Rablen 2017; Smojver 2016; Goumagias, Hristu-Varsakelis, and Assael 2018; Rahimikia et al. 2017; Mabe-Madisa 2018) |
| Precisión | (Didimo et al. 2018; Mabe-Madisa 2018; Qiu et al. 2016; Goumagias, Hristu-Varsakelis, and Assael 2018; Jupri and Sarno 2018; Höglund 2017; de Roux et al. 2018; Ceccato and Benson 2016; Piolatto and Rablen 2017; Babu and Vasavi 2017; Lismont et al. 2018; Levaggi and Menoncin 2016; Tian et al. 2016; Zhu et al. 2018; Smojver 2016; Brum and Crokidakis 2017; Hemberg et al. 2016; Rahimikia et al. 2017; Ausloos, Cerqueti, and Mir 2017; Zheng et al. 2016; Didimo et al. 2018) |
| Recall | (Didimo et al. 2018; Mabe-Madisa 2018; Qiu et al. 2016; Goumagias, Hristu-Varsakelis, and Assael 2018; Jupri and Sarno 2018; Höglund 2017; de Roux et al. 2018; Ceccato and Benson 2016; Piolatto and Rablen 2017; Babu and Vasavi 2017; Lismont et al. 2018; Levaggi and Menoncin 2016; Tian et al. 2016; Zhu et al. 2018; Smojver 2016; Brum and Crokidakis 2017; Hemberg et al. 2016; Rahimikia et al. 2017; Ausloos, Cerqueti, and Mir 2017; Zheng et al. 2016; Didimo et al. 2018) |
| F Beta Score | (Didimo et al. 2018; Levaggi and Menoncin 2016; Tian et al. 2016; Zhu et al. 2018; Smojver 2016; Rahimikia et al. 2017; Ausloos, Cerqueti, and Mir 2017; Didimo et al. 2018) |
| Curva ROC | (Didimo et al. 2018; Mabe-Madisa 2018; Tian et al. 2016; Zhu et al. 2018; Smojver 2016; Ausloos, Cerqueti, and Mir 2017; Zheng et al. 2016; Didimo et al. 2018) |

##### REFERENCIAS

Ausloos, Marcel, Roy Cerqueti, and Tariq A. Mir  
 2017 Data Science for Assessing Possible Tax Income Manipulation: The Case of Italy. Chaos, Solitons & Fractals 104: 238–256.

Babu, S Kishore, and S Vasavi  
 2017 Predictive Analytics as a Service on Tax Evasion Using Gaussian Regression Process 7: 6.

Brum, Rafael M., and Nuno Crokidakis  
 2017 Dynamics of Tax Evasion through an Epidemic-like Model. International Journal of Modern Physics C 28(02): 1750023.

Ceccato, Vania, and Michael L. Benson  
 2016 Tax Evasion in Sweden 2002–2013: Interpreting Changes in the Rot/Rut Deduction System and Predicting Future Trends. Crime, Law and Social Change 66(2): 217–232.

Didimo, Walter, Luca Giamminonni, Giuseppe Liotta, Fabrizio Montecchiani, and Daniele Pagliuca  
 2018 A Visual Analytics System to Support Tax Evasion Discovery. Decision Support Systems 110: 71–83.

Goumagias, Nikolaos D., Dimitrios Hristu-Varsakelis, and Yannis M. Assael  
 2018 Using Deep Q-Learning to Understand the Tax Evasion Behavior of Risk-Averse Firms. Expert Systems with Applications 101: 258–270.

Hemberg, Erik, Jacob Rosen, Geoff Warner, Sanith Wijesinghe, and Una-May O’Reilly  
 2016 Detecting Tax Evasion: A Co-Evolutionary Approach. Artificial Intelligence and Law 24(2): 149–182.

Höglund, Henrik  
 2017 Tax Payment Default Prediction Using Genetic Algorithm-Based Variable Selection. Expert Systems with Applications 88: 368–375.

Jupri, M., and Riyanarto Sarno  
 2018 Taxpayer Compliance Classification Using C4.5, SVM, KNN, Naive Bayes and MLP. *In* 2018 International Conference on Information and Communications Technology (ICOIACT) Pp. 297–303. Yogyakarta: IEEE. https://ieeexplore.ieee.org/document/8350710/, accessed December 18, 2018.

Levaggi, Rosella, and Francesco Menoncin  
 2016 Optimal Dynamic Tax Evasion: A Portfolio Approach. Journal of Economic Behavior & Organization 124: 115–129.

Lismont, Jasmien, Eddy Cardinaels, Liesbeth Bruynseels, et al.  
 2018 Predicting Tax Avoidance by Means of Social Network Analytics. Decision Support Systems 108: 13–24.

Mabe-Madisa, G. V.  
 2018 A Decision Tree and Naïve Bayes Algorithm for Income Tax Prediction. African Journal of Science, Technology, Innovation and Development 10(4): 401–409.

Piolatto, Amedeo, and Matthew D. Rablen  
 2017 Prospect Theory and Tax Evasion: A Reconsideration of the Yitzhaki Puzzle. Theory and Decision 82(4): 543–565.

Qiu, Junfei, Qihui Wu, Guoru Ding, Yuhua Xu, and Shuo Feng  
 2016 A Survey of Machine Learning for Big Data Processing. EURASIP Journal on Advances in Signal Processing 2016(1). https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-016-0355-x, accessed December 18, 2018.

Rahimikia, Eghbal, Shapour Mohammadi, Teymur Rahmani, and Mehdi Ghazanfari  
 2017 Detecting Corporate Tax Evasion Using a Hybrid Intelligent System: A Case Study of Iran. International Journal of Accounting Information Systems 25: 1–17.

de Roux, Daniel, Boris Perez, Andrés Moreno, Maria del Pilar Villamil, and César Figueroa  
 2018 Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach. *In* Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD ’18 Pp. 215–222. London, United Kingdom: ACM Press. http://dl.acm.org/citation.cfm?doid=3219819.3219878, accessed December 18, 2018.

Smojver, Slaven  
 2016 Inspecting Compliance to Many Rules: An Agent-Based Model. Interdisciplinary Description of Complex Systems 14(3): 277–295.

Tian, Feng, Tian Lan, Kuo-Ming Chao, et al.  
 2016 Mining Suspicious Tax Evasion Groups in Big Data. IEEE Transactions on Knowledge and Data Engineering 28(10): 2651–2664.

Zheng, Zhigang, Wei Wei, Chunming Liu, et al.  
 2016 An Effective Contrast Sequential Pattern Mining Approach to Taxpayer Behavior Analysis. World Wide Web 19(4): 633–651.

Zhu, Xulyu, Zheng Yan, Jianfei Ruan, Qinghua Zheng, and Bo Dong  
 2018 IRTED-TL: An Inter-Region Tax Evasion Detection Method Based on Transfer Learning. *In* 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) Pp. 1224–1235. New York, NY, USA: IEEE. https://ieeexplore.ieee.org/document/8456038/, accessed December 18, 2018.