

PROYECTO DE INVESTIGACIÓN DATOS INFORMATIVOS

TIPO DE CONVOCATORIA

Proyecto Interno Proyecto Semilla Proyecto Junior Proyecto Multi e Interdisciplinario

Fecha de presentación (dd/mm/aa): 01/08/2018

Título del proyecto: *(Revisar la guía para la presentación de las propuestas de los proyectos de investigación)*

Marco de trabajo para la predicción de contribuyentes deudores con alto riesgo de no pago de impuestos usando técnicas de minería de datos.

TIPOS DE INVESTIGACIÓN

Investigación básica

Investigación aplicada

DEPARTAMENTO(S) Y/O INSTITUTO:

1. Departamento de Informática y Ciencias de Computación

LÍNEA(S) DE INVESTIGACIÓN (verificable en el SAEW):

1. Computación aplicada a los Sistemas de Información ✓

RESUMEN DE INFORMACIÓN DEL DIRECTOR Y CODIRECTOR

Director					
Apellidos y nombres	No. de Cédula	HSS*	Departamento	Título de mayor nivel y mención.	No. ORCID
Hallo Carrasco María	1801087147	12	DICC	PhD	https://orcid.org/0000-0002-6718-0603

Codirector <i>(Se aplica para todos los proyectos, el codirector será a su vez colaborador)</i>					
Apellidos y nombres	No. de Cédula	HSS*	Departamento	Título de mayor nivel y mención.	No. ORCID
Marco Molina	1705613014	2	DICC	PHD	0000-0002-1055-1450



RESUMEN DE INFORMACIÓN COLABORADORES

Colaborador(es)					
Apellidos y nombres	No. de Cédula	HSS*	Departamento	Título de mayor nivel y mención.	No. ORCID

Estudiantes de Posgrado de Investigación					
Apellidos y nombres	No. de Cédula	HSS*	Programa de Posgrado	Título de mayor nivel y mención.	No. ORCID
José Ordoñez		N/A	Maestría de Investigación	Ingeniero de Sistemas	

Colaboradores Externos					
Apellidos y nombres	No. de identificación	HSS*	Institución	Título de mayor nivel y mención.	No. ORCID
Luján Mora Sergio		N/A	Universidad de Alicante	PhD	https://orcid.org/0000-0001-5000-864X

* HSS =Horas Semana Semestre: Es el número de horas que se dedica por semana a la investigación. Este número de horas se mantiene para todo el semestre

PROYECTO DE INVESTIGACIÓN

Proyecto Interno Proyecto Semilla Proyecto Junior Proyecto Multi e Inter Disciplinario

Investigación Básica

Investigación Aplicada

DEPARTAMENTO(S) Y/O INSTITUTOS:

1. Departamento de Informática y Ciencias de Computación

LINEA(S) DE INVESTIGACIÓN:

1. Computación aplicada a los Sistemas de Información

CAMPO DEL CONOCIMIENTO (Ver Anexo A: Detalle de los campos del conocimiento)

Campo amplio	Campo detallado	Campo específico
Información y Comunicación (TICS)	Información y Comunicación (TICS)	Software y análisis de aplicaciones

DISCIPLINA CIENTÍFICA (Marque X, solamente una opción)

Ciencias Naturales y Exactas	
Ingeniería y Tecnologías	x
Ciencias Médicas	
Ciencias Agrícolas	
Ciencias Sociales	
Humanidades	

OBJETIVO SOCIOECONÓMICO (Marque X, solamente una opción)

Exploración y explotación del medio terrestre	
Ambiente	
Exploración y explotación del espacio	
Transporte, telecomunicaciones y otras infraestructuras	
Energía	
Producción y tecnología industrial	
Salud	
Agricultura	
Educación	
Cultura, ocio, religión y medios de comunicación	
Sistemas políticos y sociales, estructuras y procesos	
Defensa	
Avance general del conocimiento: I+D financiada con los Fondos Generales de Universidades (FGU)	x
Avance general del conocimiento: I+D financiados con otras fuentes	

Alcance Territorial (Marque X, solamente una opción)

Institucional	Nacional	x
Parroquial	Internacional	
Cantonal	No definido	



Provincial

1 Proyecto de Investigación
Título (mínimo 10 palabras): Marco de trabajo para la predicción de contribuyentes deudores con alto riesgo de no pago de impuestos usando técnicas de minería de datos.
Resumen del proyecto (máximo 200 palabras) Uno de los procesos primarios en una administración tributaria es la Gestión de Cobranza. Los recursos que disponen las administraciones tributarias no son ilimitados, las administraciones buscan recuperar valores en etapas tempranas donde los costos de control son menores. En este sentido, las administraciones tributarias han intentado modelar el riesgo de no pago de los contribuyentes para tomar decisiones oportunamente. Para optimizar el proceso de Gestión de Cobranza y aportar en la toma de decisiones, se propone elaborar un marco de trabajo que permita el desarrollo y/o integración de componentes tecnológicos para la detección de patrones de comportamiento de los contribuyentes deudores y clasificarlos de acuerdo a su riesgo. Una vez categorizados los contribuyentes se desarrollará un modelo, usando técnicas de minería de datos, que prediga qué contribuyentes se incorporarán en la cartera vencida en los próximos años para ayudar a las administraciones tributarias en sus procesos de reducción de dicha cartera. Como caso de prueba, se trabajará con una muestra de datos de una administración tributaria del país.
Palabras clave (4-6): Gestión de cobranzas, Inteligencia de Negocios, Minería de datos.

2 Objetivos, limitaciones, hipótesis y resultados esperados de esta propuesta de investigación

2.1 Objetivos

2.1.1 Objetivo General

Desarrollar un marco de trabajo para la predicción de contribuyentes deudores con alto riesgo de no pago de impuestos usando técnicas de minería de datos.

2.1.2 Objetivos Específicos

- Realizar el estudio del estado del arte relacionado con la investigación.
- Desarrollar un marco de trabajo para el proceso de predicción de contribuyentes deudores con alto riesgo de no pago, para apoyar a la gestión de cobranzas de impuestos.
- Desarrollar un modelo de predicción de contribuyentes deudores con riesgo de no pago, para apoyar a la gestión de cobranzas tributarias usando técnicas de minería de datos.
- Publicar los resultados de la investigación.

2.2 Limitaciones (Aspectos que quedan fuera del alcance del Proyecto de Investigación)

El modelo se probará con una muestra de datos reales estadísticamente significativa y será dirigido a un tipo de contribuyentes que se considere de mayor relevancia.



El modelo apoyará a las administraciones tributarias nacionales (SRI) y se elaborará en base a datos históricos existentes.

2.3 Hipótesis

a. Los contribuyentes utilizan procesos sistemáticos no detectables en los controles realizados actualmente por las administraciones tributarias para evasión fiscal que pueden ser detectados con técnicas de minería de datos.

2.3 Detalle de los resultados esperados (con relación a los objetivos)

- Un estudio del estado del arte relacionado con la investigación.
- Un marco de trabajo para el proceso de predicción de contribuyentes deudores con riesgo de no pago para apoyo al proceso de gestión de cobranzas tributarias. (Procedimientos y arquitectura técnica).
- Un modelo de predicción de contribuyentes deudores con riesgo de no pago en la gestión de cobranzas usando técnicas de minería de datos.
- Una o varias publicaciones en congresos o revistas científicas.

3 Relevancia de la propuesta de investigación y su relación con la(s) líneas de investigación

Esta propuesta de investigación refuerza las líneas de investigación en Computación aplicada a Sistemas de Información y Machine Learning. Los resultados de la investigación se probarán con una muestra de datos reales de una administración tributaria nacional con el objetivo de contribuir a reducir la cartera vencida del estado. El director del proyecto y su codirector son profesores de las cátedras de Inteligencia de negocios, Minería de datos y Aprendizaje de máquina en la FIS por lo que los resultados de investigación serán usados como recursos de aprendizaje de las materias.

4 Impacto de la investigación

4.1 Impacto Social

Actualmente la gestión de cobranzas en las administraciones tributarias no permite detectar a tiempo contribuyentes deudores con riesgo de no pago, por lo que el trabajo de investigación contribuirá a solucionar el problema detectado, contribuyendo a mejorar la liquidez de las instituciones estatales para cumplir con los compromisos de pago.

4.2 Impacto Económico

El marco de trabajo y modelo de predicción a desarrollarse contribuirá a disminuir la cartera vencida del estado con lo que mejorará ostensiblemente la recaudación fiscal en beneficio de la ejecución de proyectos nacionales.

4.3 Impacto Político

La investigación contribuirá a mejorar la relación Universidad –Estado mediante la ejecución conjunta de proyectos que permitan resolver problemas nacionales.

4.4 Impacto Científico

El desarrollo del marco de trabajo y modelo de predicción permitirá investigar el estado del arte de nuevas tecnologías de minería de datos y proponer modelos que se adapten a la realidad de las administraciones tributarias.

5 Productos esperados

Tipo de Producto:	Marcar con una "X"
a. Publicaciones científicas y/o patente (obligatorio);	x
b. Disertación a la comunidad politécnica;	x



c. Trabajo de titulación de acuerdo a lo que establece el Reglamento de Régimen Académico y la Normativa Interna de la EPN;	x

6 Descripción, metodología y diseño del proyecto

6.1 Descripción, metodología y diseño del proyecto (Máximo dos carillas)

El proyecto comprende el desarrollo de un marco de trabajo para integrar y o desarrollar tecnologías y un modelo de predicción de contribuyentes deudores con riesgo de no pago, en la gestión de cobranzas usando técnicas de minería de datos. Existen investigaciones que han contribuido a demostrar que el uso de herramientas de minería de datos permite agregar valor al trabajo de recaudación de impuestos [1,2,3,4,5]. Por lo que el desarrollo del trabajo propuesto contribuirá a cumplir con el objetivo planteado.

El marco de trabajo será desarrollado de forma iterativa e incremental y será probado con datos de la administración tributaria nacional. Los resultados serán publicados al menos en un artículo de una revista indexada y o en un congreso.

Para la revisión de literatura se seguirá una adaptación de la metodología de Kitchenham [6,7,8]. El resultado buscado es el estudio del estado del arte de la aplicación de técnicas de minería de datos en administraciones tributarias.

Para el desarrollo del modelo de predicción, se utilizará la metodología CRISP-DM [9] para minería de datos. A través de esta metodología se pretende el desarrollo de un modelo de predicción de contribuyentes deudores, con riesgo de no pago, en la gestión de cobranzas usando técnicas de minería de datos. La metodología utiliza las siguientes etapas:

- Revisión bibliográfica.
- Entendimiento del negocio: Permitirá entender las problemáticas de la Gestión de Cobranzas y establecer definiciones del proyecto de minería de datos para las administraciones tributarias y como caso de prueba, el SRI.
- Entendimiento de los datos: Incluye la obtención de datos a investigarse, exploración, verificación y reporte de su calidad.
- Revisión de tecnologías de apoyo.
- Preparación de los datos: Permite escoger los datos necesarios para el análisis de acuerdo a criterios seleccionados en la fase anterior. Estos datos alimentarán las herramientas de modelado.
- Modelado: Seleccionar la técnica de modelado, probar la calidad del modelo seleccionado y validarlo. Usando herramientas de modelado, se construirá el modelo. Finalmente se validará el contexto de la Gestión de Cobranzas.
- Evaluación de resultados: Valorar si los modelos responden a los objetivos propuestos en la investigación.
- Despliegue: Presentación de los resultados propuestos en la presente investigación.
- Difusión de resultados

Adicionalmente se realizará la preparación y envío de publicaciones.

El proyecto contempla una estancia de investigación en la universidad de Alicante por dos meses sin costo para la EPN en viáticos o pasajes.

Referencias Bibliográficas

- [1] DeBarr, D., & Eyster-Walker, Z. (2006). Closing the gap: automated screening of tax returns to identify egregious tax shelters. *ACM SIGKDD Explorations Newsletter*, 8(1), 11-16.
- [2] Pijnenburg, M. G. F., Kowalczyk, W. J., & van der Hel-van Dijk, E. C. J. M. (2017). A Roadmap for Analytics in Tax payer Supervision. *Electronic Journal of e-Government*, 15, 14.



- [3] Rad, M. S., & Shahbahrami, A. (2016, December). Detecting high risk taxpayers using data mining techniques. In *Signal Processing and Intelligent Systems (ICSPIS), International Conference of* (pp. 1-5). IEEE.
- [4] Predictive Analytics in the Public Sector: Using Data Mining to Assist Better Target Selection for Audit.
- [5] Tian, F., Wan, X., Lan, T., Qi, T., Zheng, Q., Chao, K. M., & Li, L. (2016). Find the most suspicious tax evasion groups from a taxpayer interest interacted network. In *SMC* (pp. 3624-3629).
- [6] Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – a systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7 – 15, 2009.
- [7] Sander, L., & Kitcher, H. (2006). Systematic and other reviews: terms and definitions used by UK organisations and selected databases Systematic review and Delphi survey. *National Institute for Health and Clinical Excellence*.
- [8] Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, 14(3), 207-222.
- [9] R. Wirth, "Crisp-dm: Towards a standard process model for datamining," in *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 2000, pp. 29–39.

7	Infraestructura, equipos y fondos adicionales.
----------	---

7.1 Infraestructura y equipos

- Indicar la infraestructura y equipos **disponibles** para la ejecución del proyecto, con la ubicación actual de los mismos

Infraestructura	Equipos	
Laboratorio BIlab	Nombre del Equipo	Ubicación del Equipo
	Altex	Laboratorio BILab, Departamento DICC
	Impresora HP M127fn	Laboratorio BILab,

7.2 Breve justificación del equipo requerido

Se usará equipamiento computacional y de impresión existente.

7.3 Fondos Adicionales

Aportes en tiempo del colaborador externo.



VICERRECTORADO DE INVESTIGACIÓN Y PROYECCIÓN SOCIAL
UNIDAD DE INVESTIGACIÓN
PRESUPUESTO PROYECTOS DE INVESTIGACIÓN



AÑO 1

Título del proyecto

Marco de trabajo para la predicción de contribuyentes deudores con alto riesgo de no pago de impuestos usando técnicas de minería de datos

Lista de Items		Cantidad	Unidad	Precio Unitario Referencial	Precio Total Referencial	Precio Unitario Referencial +Aporte IESS	Precio Total Referencial con IVA + Aporte del IESS
1	Contratación de servicios personales por contrato						
1.1	Ayudantes de investigación	7	mes	\$ 125,00	\$ 875,00	\$ 147,00	\$ 1.029,00
1.2	Prestación de servicios profesionales (Homologado Escala de remuneración de servidores publicos)		mes		\$ -	\$ -	\$ -
Subtotal 1				\$ 125,00	\$ 875,00	\$ 147,00	\$ 1.029,00
Lista de Items		Cantidad	Unidad	Precio Unitario Referencial sin IVA	Precio Total Referencial sin IVA	Precio Unitario Referencial con IVA	Precio Total Referencial con IVA
2	Maquinaria y equipo especializado						
2.1	Item 1 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
2.2	Item 2 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
2.3	Item 3 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
2.4	Item 4 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
Subtotal 2				\$ -	\$ -	\$ -	\$ -
3	Equipo Informático						
3.1	Item 1 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
3.2	Item 2 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
3.3	Item 3 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
3.4	Item 4 (Detallar nombre de la maquinaria y equipos solicitado)			\$ -	\$ -	\$ -	\$ -
Subtotal 3				\$ -	\$ -	\$ -	\$ -
4	Insumos y reactivos						
4.1	Item 1 (Detallar nombre de los insumos y reactivos)			\$ -	\$ -	\$ -	\$ -
4.2	Item 2 (Detallar nombre de los insumos y reactivos)			\$ -	\$ -	\$ -	\$ -
4.3	Item 3 (Detallar nombre de los insumos y reactivos)			\$ -	\$ -	\$ -	\$ -
4.4	Item 4 (Detallar nombre de los insumos y reactivos)			\$ -	\$ -	\$ -	\$ -
Subtotal 4				\$ -	\$ -	\$ -	\$ -
5	Literatura especializada						
5.1	Item 1 (Detallar nombre del libro)			\$ -	\$ -	\$ -	\$ -
5.2	Item 2 (Detallar nombre del libro)			\$ -	\$ -	\$ -	\$ -
5.3	Item 3 (Detallar nombre del libro)			\$ -	\$ -	\$ -	\$ -
5.4	Item 4 (Detallar nombre del libro)			\$ -	\$ -	\$ -	\$ -
Subtotal 5				\$ -	\$ -	\$ -	\$ -
6	Salidas de campo y de muestreo						
6.1	Pasajes al interior			\$ -	\$ -	\$ -	\$ -
6.2	Viaticos y subsistencias al interior			\$ -	\$ -	\$ -	\$ -
Subtotal 6				\$ -	\$ -	\$ -	\$ -
7	Ponencias nacionales, capacitaciones y/o visitas técnicas						
7.1	Pasajes al interior			\$ -	\$ -	\$ -	\$ -
7.2	Viaticos y subsistencias al interior			\$ -	\$ -	\$ -	\$ -
Subtotal 7				\$ -	\$ -	\$ -	\$ -
8	Ponencias en el exterior, capacitaciones, y/o visitas técnicas						
8.1	Pasajes al exterior			\$ 2.000,00	\$ 2.000,00	\$ 2.240,00	\$ 2.240,00
8.2	Viaticos al exterior			\$ 1.000,00	\$ 1.000,00	\$ 1.000,00	\$ 1.000,00
Subtotal 8				\$ 3.000,00	\$ 3.000,00	\$ 3.240,00	\$ 3.240,00
9	Pago de inscripciones						
9.1	Pago de inscripciones al interior			-	\$ -	\$ -	\$ -
9.2	Pago de inscripciones al exterior			-	\$ -	\$ -	\$ 700,00
Subtotal 9				\$ -	\$ -	\$ -	\$ 700,00
10	Pago de publicaciones y patentes						
10.1	Pago de publicaciones			\$ -	\$ -	\$ -	\$ -
10.2	Pago de publicaciones al exterior			\$ -	\$ -	\$ -	\$ -
10.2	Pago de patentes			\$ -	\$ -	\$ -	\$ -
Subtotal 10				\$ -	\$ -	\$ -	\$ -
TOTAL					\$ 3.875,00		\$ 4.969,00



VICERRECTORADO DE INVESTIGACIÓN Y PROYECCIÓN SOCIAL
UNIDAD DE INVESTIGACIÓN
PRESUPUESTO PROYECTOS DE INVESTIGACIÓN



Título del proyecto

Marco de trabajo para la predicción de contribuyentes deudores con alto riesgo de no pago de impuestos usando técnicas de minería de datos

Presupuesto consolidado sin IVA

AÑO	Contratación de servicios personales por contrato	Maquinaria y equipo especializado	Equipo informático	Insumos y reactivos	Literatura especializada	Salidas de campo y de muestreo	Ponencias nacionales, capacitaciones y/o visitas técnicas	Ponencias en el exterior, capacitaciones, y/o visitas técnicas	Pago de inscripciones	Pago de publicaciones y patentes	Total sin IVA
1	\$ 875,00	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ 3.000,00	\$ -	\$ -	\$ 3.875,00
2	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -
3	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -
TOTAL	\$ 875,00	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ 3.000,00	\$ -	\$ -	\$ 3.875,00

Presupuesto consolidado con IVA

AÑO	Contratación de servicios personales por contrato	Maquinaria y equipo especializado	Equipo informático	Insumos y reactivos	Literatura especializada	Salidas de campo y de muestreo	Ponencias nacionales, capacitaciones y/o visitas técnicas	Ponencias en el exterior, capacitaciones, y/o visitas técnicas	Pago de inscripciones	Pago de publicaciones y patentes	Total con IVA
1	\$ 1.029,00	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ 3.240,00	\$ 700,00	\$ -	\$ 4.969,00
2	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -
3	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -
TOTAL	\$ 1.029,00	\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	\$ 3.240,00	\$ 700,00	\$ -	\$ 4.969,00

DECLARACIÓN FINAL

TIPO DE PROYECTO

Proyecto Interno Proyecto Semilla Proyecto Junior Proyecto Multi e Interdisciplinario

TIPO DE INVESTIGACIÓN

Investigación básica Investigación aplicada

TÍTULO DEL PROYECTO

Marco de trabajo para la predicción de contribuyentes deudores con alto riesgo de no pago de impuestos usando técnicas de minería de datos

DECLARACIÓN DEL DIRECTOR DEL PROYECTO

El equipo de investigadores, representado por el Director del Proyecto declara lo siguiente:

- Que el presente proyecto es una obra original de este equipo de investigadores y por tanto, asumimos la completa responsabilidad legal en caso de que un tercero alegue la titularidad de los derechos intelectuales del proyecto, exonerando a la EPN de cualquier acción legal que se derive por esta causa.
- Que el presente proyecto no ha sido presentado en ninguna convocatoria de otra institución pública o privada solicitando el financiamiento total del presupuesto. El incumplimiento será causal para que la propuesta sea descalificada de la convocatoria de la EPN.
- Que todos los bienes adquiridos en el proyecto permanecerán bajo la custodia y responsabilidad del director de proyecto.
- Que aceptamos que si el proyecto genera algún producto o procedimiento susceptible de obtener de derechos de propiedad intelectual, de los cuales se deriven beneficios, éstos serán compartidos entre los investigadores y la EPN.
- Que aceptamos conocer y cumplir con la normativa vigente para la gestión de proyectos de investigación.



Firma del Director del Proyecto
Nombre: Dra. María Hallo
C.I.:1801087147

DECLARACIÓN DEL JEFE DE DEPARTAMENTO

Esta propuesta ha sido avalada por el Consejo del Departamento de Informática y Ciencias de la Computación, en sesión del día 06 de agosto de 2018 mediante resolución No. 138.033.06-08-2018.

Las instalaciones, incluyendo personal, edificios, equipo y recursos financieros están a disposición del proponente y sus colaboradores de acuerdo con las especificaciones que se encuentran en esta propuesta.



Firma del Jefe del Departamento
Nombre: Dra. Myriam Peñafiel Aguilar
C.I.: 1705828711



*Se debe adjuntar el acta en el que conste la resolución que avala la propuesta de proyecto