

Tipo de artículo: Artículo original
Temática: Inteligencia Artificial
Recibido:17/06/2016 | Aceptado:10/10/2016

Método para el aseguramiento de ingresos en entornos de desarrollo de software

Revenue Assurance's Method in Software Development's Environments

Gilberto Fernando Castro Aguilar^{1*}, Iliana Pérez Pupo², Pedro Y. Piñero Pérez², Roberto García Vacacela¹

¹Universidad Católica de Santiago de Guayaquil, Ecuador. gfercastro@gmail.com, roberto.garcia@cu.ucsg.edu.ec

²Laboratorio de Investigaciones de Gestión de Proyecto. Universidad de las Ciencias Informáticas. Carretera a San Antonio de los Baños Km, 2 ½. La Lisa, La Habana, Cuba, cp.:19370. iperez@uci.cu, ppp@uci.cu

Autor para correspondencia: gfercastro@gmail.com

Resumen

En los entornos orientados a proyectos de desarrollo de software ocurren numerosos errores tanto en la planificación como en el control y seguimiento de los mismos que están provocando pérdida de ingresos en las organizaciones. Es un hecho solo aproximadamente el 38% de los proyectos de tecnología de la información son satisfactorios, el resto son renegociados o cancelados. Estas dificultades provocan pérdidas de ingresos en los proyectos y las organizaciones que lo desarrollan. En este trabajo se propone un método para el aseguramiento de ingresos en organizaciones orientadas a proyectos que combina técnicas de la gestión de proyectos con técnicas de minería de datos y *soft computing*. Entre las técnicas de gestión de proyectos empleadas se encuentran: técnicas de gestión de riesgos, las técnicas asociadas a la planificación de tiempo, planificación de los costos, la gestión del alcance y la gestión de las adquisiciones. Entre las técnicas de minería de datos y *soft computing* que se combinan se encuentran: las técnicas para la minería de datos anómalos, el aprendizaje automático y la computación con palabras. Se proponen de esta forma un método novedoso para el tratamiento de la incertidumbre de la información y la imprecisión de los datos en los procesos de planificación de proyectos. En la validación de los resultados se emplean técnicas de validación cruzada para la comparación de varios métodos, a partir de su aplicación en bases de datos del repositorio de datos para investigaciones del Laboratorio de Investigaciones en Gestión de Proyectos. Los resultados del trabajo aportaron que el algoritmo CLARA reportó mejores resultados que el KMeans y que el uso de la computación con palabras en la evaluación de los riesgos aporta mayor robustez en la propuesta.

Palabras clave: Aseguramiento de ingresos; gestión de proyectos; minería de datos anómalos.

Abstract

In software projects usually many defects and human errors are introduced. Some of these errors have to do with the planning processes whereas others are associated to control and making decisions processes. Frequently revenue assurance processes are affected for these errors. In fact, just 38% of projects are considered successfully whereas the rest are challenged or canceled. In this work, authors proposed a method, for revenue assurance in project management organizations based on combination of project management techniques with data mining and soft computing techniques. Different project management techniques, such as: risk management techniques, scheduling techniques, scope techniques and cost techniques, are combined with computer sciences techniques. In order, to manage imprecision and uncertainty on data, authors proposed the use soft computing techniques such as computing with words and fuzzy machine learning techniques. Authors applied the method in project management data sets published in repository of research of Project Management Laboratory. Authors applied cross-validation techniques by comparing different methods, CLARA algorithm report better results than Kmeans algorithm.

Keywords: Revenue assurance; Software Projects; Soft computing; Project management.

Introducción

A finales de la década de los 70' surge el aseguramiento de ingresos con el objetivo de proteger y recuperar los recursos financieros de las empresas de telecomunicaciones. De hecho, algunos autores como Massyn conceptualizan el aseguramiento de ingresos “como una disciplina emergente asociada a las empresas de telecomunicaciones con alto nivel de dinamismo que requiere enfoques proactivos para su desarrollo” (Massyn, 2010). No obstante, al origen de la disciplina, debido a su aplicabilidad, esta se ha extendido rápidamente abarcando así las organizaciones orientadas a la gestión de proyectos de software. En este nuevo contexto han surgido definiciones más generales para describir el objeto de estudio de esta disciplina. Entre ellos se destaca Khan quien plantea que “el aseguramiento de ingresos es el conjunto de actividades que son aplicadas para asegurar que los procesos del negocio, la estructura organizacional, los controles y los sistemas de información, relacionadas con el ciclo de ingresos de las organizaciones, trabajen juntos con efectividad” (Khan, 2014). Acosta, por su parte plantea “consiste en buscar, identificar y eliminar las causas técnicas y estructurales que dan origen a las fugas de ingresos dirigida a dos dimensiones la maximización de los ingresos y la minimización de los costos” (Acosta, 2008).

Con el desarrollo de la disciplina de aseguramiento de ingresos surgen y se fortalecen los espacios TMForum (TM Forum, 2015) y la Asociación Global de Profesionales de Aseguramiento de Ingresos GRAPA (Mattison, 2009) como espacios para la formación de especialistas y a la creación de estándares. TMForum (TM Forum, 2015) ha desarrollado

los modelos de evaluación TR131 y GB941, basados en 5 niveles: inicial, repetible, definido, manejado y optimizado que permiten evaluar la madurez de organizaciones en el aseguramiento de ingresos. Por su parte GRAPA, ha desarrollado un estándar con alto nivel de reconocimiento que propone método a seguir para aplicar las técnicas de esta disciplina.

En el caso particular de los proyectos de tecnologías de la información la disciplina de aseguramiento de ingresos adquiere paulatinamente mayor relevancia. Es un hecho que en este sector la media histórica de proyectos satisfactorios es 30,7 % mientras que los proyectos renegociados son el 47,3 % y los cancelados el 22 %, ver Figura 1, (The Standish Group International, 2014). Estas cifras implican que enormes presupuestos son afectados cada año por concepto de errores en la planificación o el control y seguimiento de los proyectos, con un impacto económico y social.

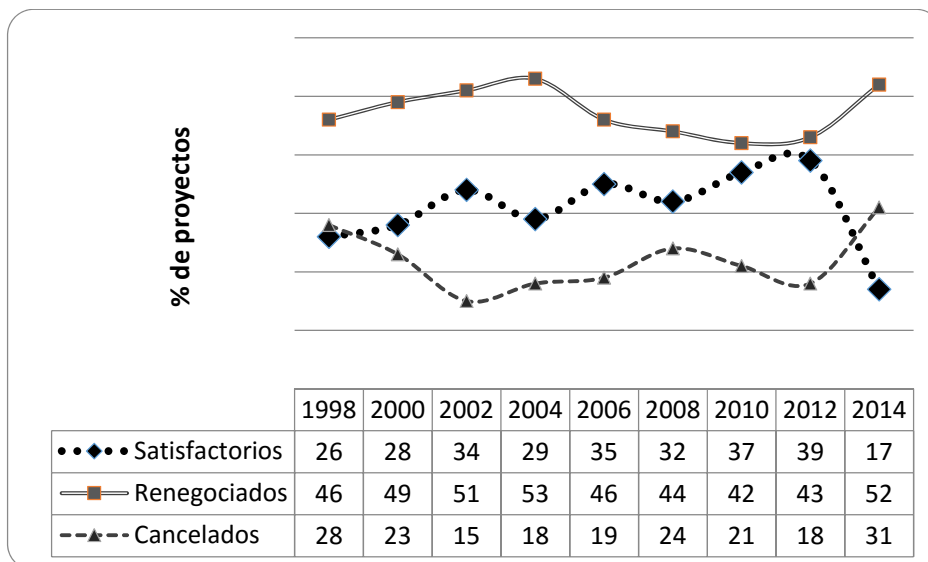


Figura 1. Porcientos de éxito de los proyectos.

El gran volumen de información que se gestiona en las organizaciones orientadas a proyectos de software asociada a: la gestión financiera, los planes de proyectos, a la gestión de recursos entre otras, ha motivado a los empresarios e investigadores a desarrollar nuevos métodos para el aseguramiento de ingresos en dichas organizaciones. Según (Mattison, 2009) entre las técnicas computacionales más empleadas en el aseguramiento de ingresos se encuentran: el análisis de riesgos, análisis de intercambio, análisis de procesos, análisis de sistemas y el análisis estadístico. Pero las técnicas existen aún son insuficientes porque no manejan adecuadamente problemas con la imprecisión y la incertidumbre de los datos (Singh y Upadhyaya, 2012).

El objetivo de este trabajo es un nuevo método para el aseguramiento de ingresos en entornos orientados al desarrollo de proyectos de software. El método propuesto se basa en la combinación de técnicas tradicionales de la gestión de proyectos con técnicas de la computación emergente para la resolución de este problema.

En la sección del trabajo metodología computacional se propone el método de investigación utilizado, se presenta el análisis de técnicas de aseguramiento de ingresos existentes las ventajas y limitaciones fundamentales. Luego se presenta el método propuesto y los resultados de su aplicación, Finalmente se presentan las conclusiones del trabajo.

Metodología computacional

En esta sección los autores proponen un nuevo método para el aseguramiento de ingresos para ello siguieron los siguientes pasos.

1. Conceptualización de los principales problemas de aseguramiento de ingresos asociados a los proyectos de desarrollo de software.
2. Formalización de un método para el aseguramiento de ingresos en organizaciones orientadas a proyectos de software.
3. Experimentación con la base de datos de evaluación de proyectos y planificación de tareas y recursos del repositorio de investigaciones en gestión de proyectos.
4. Aplicación de la propuesta en un caso de estudio en Ecuador.
5. Análisis de los resultados.

Conceptualización de los principales problemas de aseguramiento de ingresos en los proyectos de desarrollo de software.

Según GRAPA, entre las principales causas que provocan pérdida de ingresos en las telecomunicaciones se encuentran la formación del personal, la integración de los procesos en la organización y adopción de estándares ver Figura 2.

Nótese como los factores asociados a la formación del personal juegan un importante rol en este entorno mientras que factores asociados a los presupuestos o la cantidad del personal tienen menos relevancia. Esta situación es similar en las organizaciones orientadas a proyectos, pero con la diferencia que en estos otros entornos los costos, los presupuestos y la planificación ganan en relevancia. De hecho, entre las principales causas de fracaso de los proyectos de software

se encuentran errores en la planificación del tiempo, del alcance, del presupuesto y en la planificación de las adquisiciones. También se señalan con mucha fuerza los errores en el control y seguimiento de los proyectos.

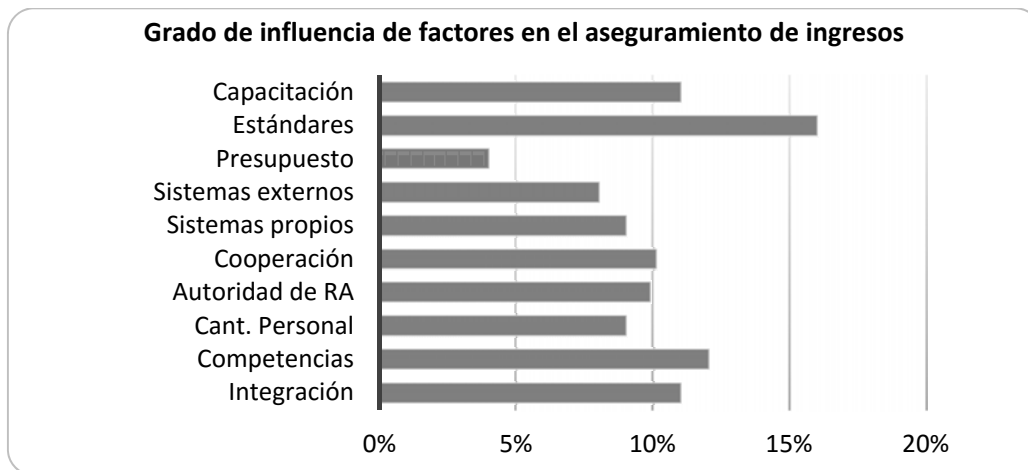


Figura 2. Impacto de las principales necesidades del Aseguramiento de Ingresos (TM Forum, 2015).

En el caso particular de los proyectos de software se identifican un conjunto de situaciones en las planificaciones que cuando aparecen acarrear errores y fallos en los proyectos:

- Tareas que no tengan la cantidad adecuada de recursos con las competencias esperadas.
- Tareas con sobreestimación del tiempo de duración o de los costos asociados.
- Cronogramas con errores que no respetan las relaciones de precedencia lógica entre las tareas o que no cubren completamente los objetivos previstos en el proyecto.
- La sobrecarga de recursos humanos y no humanos que generalmente es fuente de atrasos y conflictos.

Para detectar las fallas en los procesos de aseguramiento de ingresos y corregir los defectos se han desarrollado múltiples técnicas dentro de la minería de datos anómalos, véase algunas de las técnicas más empleadas en la Tabla 1.

Tabla 1. Técnicas de minería de datos anómalos, más empleadas en los procesos de aseguramiento de ingresos (Mattison, 2009).

Procesos de aseguramiento de ingresos	Técnicas de la minería de datos anómalos empleadas en el aseguramiento de ingresos
Proceso de análisis de riesgos	Muestreo, Análisis de grupo, Análisis de distribución, Análisis tendencia central, Regresión
Proceso de detección temprana	Muestreo, Análisis de grupo, Chaid/Cart, RNA
Proceso de diseño de controles	Muestreo, Chaid/Cart, Análisis de distribución, Análisis tendencia central
Proceso de análisis de causa raíz	Análisis de grupo, Análisis de distribución, Análisis tendenciacentral
Proceso de pronóstico	Muestreo, Análisis de grupo, Chaid/Cart, RNA y Regresión

La minería de datos anómalos ha sido una de las disciplinas más usadas en el aseguramiento de ingresos por su capacidad para el descubrimiento de patrones en los datos con comportamientos que se separan significativamente del comportamiento normal de los mismos. Existen diferentes definiciones de datos anómalos, en este trabajo nos basaremos en la definición de (Aggarwal, 2013) quien define que un dato anómalo es “una observación que se desvía mucho del resto de las observaciones, apareciendo como una observación sospechosa que pudo ser generada por mecanismos diferentes al resto de los datos”.

Pero la minería de datos anómalos incluye muchas técnicas que pueden ser muy útiles para el aseguramiento de ingresos. Algunos autores agrupan las mismas en técnicas no supervisadas, técnicas semi-supervisadas y técnicas supervisadas siendo este mismo enfoque el que se adopta en este trabajo, ver Figura 3.

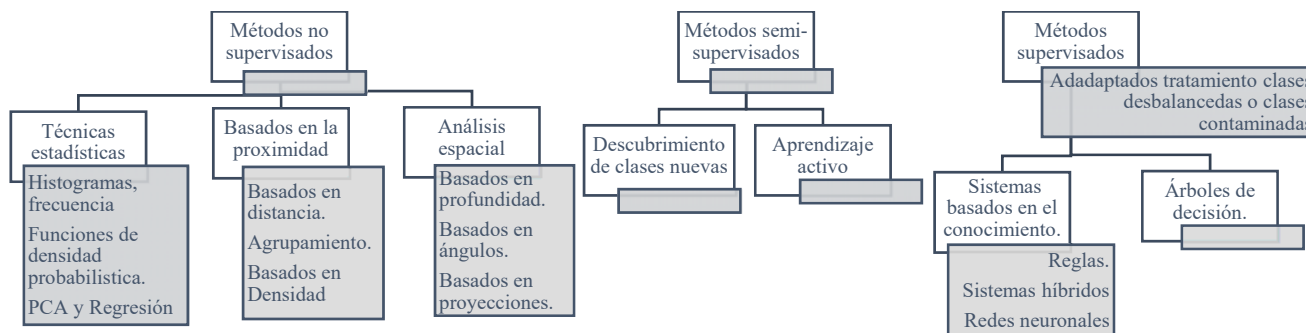


Figura 3. Organización de las técnicas de minería de datos anómalos en tres categorías.

Como ejemplo de aplicaciones de las técnicas no supervisadas para la detección de datos anómalos se presentan ejemplos de técnicas de agrupamientos en otros (Chandola, et. al., 2012; Pividori, et. al., 2015; Vijendra y Pathak, 2013; Shivani, et. al., 2015; Zhang, et. al, 2014; Sathishkumar y Thangavel, 2015), métodos basados en distancias como (Gupta, et. al., 2014; Kuna, et. al., 2014; Shpigelman, 2014; Gogoi, 2012) modelos basados en la densidad, (Zimek, et. al., 2012; Li, et. al., 2014; Deneshkumar, et. al., 2014), modelos estadísticos de basados en la distribución (Zhang, et. al, 2014). Como ejemplo de los métodos no supervisados se presentan en la bibliografía los siguientes ejemplos (Gilberto, et. al., 2015).

A pesar de la importancia del aseguramiento de ingresos y de que se han aplicado técnicas de la computación emergente a este problema, aún quedan muchas líneas abiertas de investigación, algunas de las cuales han sido formuladas por TMForum y GRAPA entre ellas se encuentran:

- La necesidad de desarrollar mayor cantidad de métodos proactivos y activos para el aseguramiento de ingresos, que disminuyan el tiempo de detección de fugas y adaptados a las peculiaridades de cada entorno específico.

- La introducción de técnicas novedosas de minería de datos anómalos y *big data* adaptables al dinamismo de las nuevas tecnologías y garantizar una exitosa aplicación de técnicas de aseguramiento de ingresos.

En particular en los entornos orientados a proyectos se identifica la combinación de técnicas tradicionales de gestión de proyectos con la minería de datos como una vía para ayudar a resolver el problema del aseguramiento de ingresos.

Se requiere el diseño de técnicas adaptadas a las situaciones que presentan los proyectos de desarrollo de software y que combinan las ventajas de diversas tecnologías.

Método propuesto para el aseguramiento de activos

En este trabajo se propone un método para el aseguramiento de ingresos aplicable en organizaciones orientadas a proyectos de software. Este método se inspira en las metodologías de análisis de datos y combina técnicas de gestión de proyectos con técnicas de minería de datos anómalos, formando una simbiosis a la medida del problema que se quiere resolver. Los pasos del método propuesto se relacionan a continuación.

- Paso 1. Comprensión del negocio y diagnóstico: tiene dos objetivos primero comprender los procesos de la organización para definir una taxonomía que ayude a identificar las situaciones que afecten los ingresos, como posibles causas de fallos, fraudes o fugas. En un segundo momento se centra en clasificar a los interesados teniendo en cuenta el impacto e interés en el proceso del aseguramiento de ingresos de la organización.
- Paso 2. Comprensión de los datos: se centra en la comprensión de las fuentes de datos, como están distribuidos los datos, la existencia de valores ausentes y en general la distribución de la información.
- Paso 3. Gestión de riesgos con un enfoque proactivo: se basa en la aplicación de técnicas de gestión de riesgos para la planificación y la evaluación cualitativa de los riesgos basada en modelos de computación con palabras.
- Paso 4. Pre-procesamiento de datos, con un enfoque reactivo: en este paso se agrupan las diferentes actividades relacionadas con la eliminación de posibles ruidos que afecten la detección de las fallas y los verdaderos datos anómalos. Se ejecutan actividades de limpieza, estandarización y la selección de los atributos.
- Paso 5. Aplicación y modelación de algoritmos de análisis reactivo de los datos: se centra en el diseño y aplicación de algoritmos para la detección de datos anómalos que reflejen pérdidas de ingresos. Generalmente estas situaciones anómalas son provocadas por fraudes o fallas de operación que son los factores que se trata de identificar y cuantificar su impacto.
- Paso 6. Evaluación de los resultados, estimación de impacto: en este paso se procede a verificar que los registros identificados sean realmente datos anómalos, y a estimar el impacto de los resultados.

Paso 7. Toma de decisiones e implantación: los resultados del análisis y la propuesta son finalmente recogidos en reportes que son analizados y donde se reflejan las principales decisiones tomadas en el análisis. Además, se procede a implantar los algoritmos y métodos para garantizar la sostenibilidad en el tiempo de la organización.

En el paso 2 se emplean técnicas de estadística descriptiva combinadas con las facilidades de los sistemas de información y el lenguaje SQL para la recuperación de datos.

En el paso 3 para la gestión de riesgos se propone la aplicación de un método que toma como base la guía del PMBoK pero que mejora los procesos de análisis cualitativo, el control y el seguimiento de los riesgos identificados. En particular se manejan tres criterios de evaluación de riesgos: probabilidad de ocurrencia, el impacto y la facilidad de detección a diferencia del PMBoK que propone solo los dos primeros. Además, se proponen técnicas de computación con palabras para el análisis cualitativo de los riesgos. También, se introducen facilidades en el sistema de información que relaciona los riesgos con las tareas en el cronograma permitiendo el cálculo de indicadores para el control.

Las técnicas de computación con palabras fueron aplicadas en los siguientes 3 pasos:

- I. Se define un conjunto básico de términos lingüísticos (*LBTL*) para la evaluación de los criterios con los siguientes términos. $LBTL = \{Ninguno, Muy\ bajo, Bajo, Medio, Alto, Muy\ alto, Perfecto\}$.
- II. Los expertos evalúan cada criterio usando alguno de los términos lingüísticos ver Tabla 2.

Tabla 2. Estructura de la evaluación de los expertos.

Criterios	Expertos		
	e_1	...	e_m
c_1	x_1^{11}	...	x_1^{1m}
\vdots	\vdots	\vdots	\vdots
c_p	x_1^{p1}	...	x_1^{pm}

- III. Se transforma la preferencia de los expertos en conjuntos borrosos considerando la variable lingüística de la Figura 4.
- IV. Siguiendo el modelo 2-tuplas de computación con palabras se agregan las evaluaciones de los expertos consolidando las mismas por cada criterio a evaluar. Ver ecuación (1).

$$x^e(x) = \left(\frac{1}{n} \sum_{i=1}^n -1((s_i, \alpha_i)) \right) = \left(\frac{1}{n} \sum_{i=1}^n \beta_i \right) \quad (1)$$

- V. Consolidar los resultados y dar la respuesta.

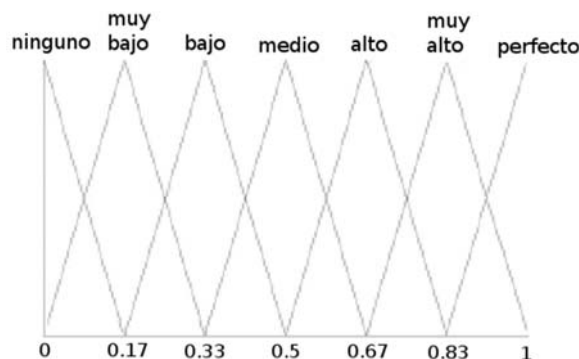


Figura 4. Variables lingüísticas usadas por los expertos para evaluar la propuesta.

En el paso 5 se introduce un método combinado para la detección de datos anómalos basado en una estrategia de tratamiento independiente de los datos con múltiples algoritmos y luego la unión de los resultados. Este algoritmo efectúa varias iteraciones y en cada una de ellas aplica un algoritmo especializado en la detección de un tipo de situación. Una vez ejecutados todos los algoritmos se combinan los resultados encontrados y se pasa al siguiente paso.

Para la evaluación de los resultados y la estimación de impacto en el paso 6 el método propone utilizar algunas de las siguientes técnicas: estimación por analogía, estimación de por tres valores, estimación paramétrica, juicio de expertos o análisis de los costos del aseguramiento de ingresos. Se explican a continuación algunas de las técnicas propuestas.

La estimación de costos por analogía: utiliza información histórica basada en la experiencia de proyectos anteriores, y el impacto de determinados factores en los mismos. Este método se basa en el uso de los sistemas basados en el conocimiento y recupera los casos semejantes a la situación actual. Posteriormente genera una posible respuesta a partir de la adaptación del nuevo caso considerando los valores de los casos similares recuperados.

Estimación por tres valores se basa en lograr estimar tres valores: costo más probable (M) basado en una evaluación realista del experto, costo optimista (O) tomando como base el mejor escenario posible con el menor impacto posible y el costo pesimista (P) basado en el análisis del peor escenario o sea el mayor impacto posible. Luego se procede a calcular el valor estimado como muestra la ecuación (1).

$$ce = \frac{O + 4M + P}{6} \quad (1)$$

La estimación paramétrica utiliza una relación estadística entre los datos históricos y otras variables para calcular una estimación del impacto económico. Esta técnica está soportada por un sistema de recomendaciones basado en el contenido y centrado en la estrategia de recuperar ítems buenos. La estimación paramétrica de costos puede aplicarse en conjunto con otros métodos de estimación.

Para el paso 7 de toma de decisiones el método propuesto propone el uso de sistemas de información para la generación de los reportes y un sistema de recomendaciones para ayudar al usuario en la toma de decisiones. Cada dato anómalo identificado será enmarcado en algunas de las situaciones que afectan los ingresos, el sistema de recomendaciones es un sistema basado en contenido y que devuelve ítems buenos, siendo cada uno de estos ítems buenos las medidas que pueden ser tomadas ante la situación. Finalmente, usuario selecciona aquellas medidas que considere más apropiadas.

Resultados y discusión

La validación de este método es compleja porque no se reportan en la bibliografía métodos similares que abarquen el proceso completo de aseguramiento de ingresos en organizaciones orientada a proyectos de software. Por tanto, se validan que los algoritmos empleados en cada una de las componentes principales del método propuesto son las más adecuadas. Se relacionan a continuación dos de las pruebas experimentales diseñadas:

- Experimento de comparación entre diferentes métodos de identificación y evaluación de riesgos.
- Experimento de comparación entre diferentes métodos de detección de tareas con costos sobregirados o muy por debajo del valor previsto.

Experimento de comparación entre diferentes métodos de identificación y evaluación de riesgos, ver Tabla 3.

Como se puede ver entre las ventajas del método propuesto está la facilidad para la interpretabilidad de los resultados al expresarse tanto la evaluación de los expertos como los resultados finales en palabras. El método propuesto además permite la valoración simultánea de múltiples expertos y como se puede ver los resultados de los dos métodos son diferentes. Ninguno de los riesgos evaluados con la evaluación tradicional PMBok, tuvieron exposición alto a pesar de que la probabilidad o el impacto de algunos, era elevado. Además, el orden final (*ord*) luego de la evaluación que es importante para la respuesta a los riesgos fue también diferente, siendo superior el método propuesto.

El experimento de comparación entre diferentes métodos de detección de tareas con costos sobregirados o muy por debajo del valor previsto. Para este experimento se empleó la base de datos de “Asignación de Recursos y Tiempo” del Laboratorio de Investigaciones en Gestión de Proyectos que contiene 9315 tareas agrupadas en 88 proyectos. Se generaron de forma aleatoria 12 particiones del espacio de búsqueda cada una con el 5% de datos anómalos. Se compararon los métodos de agrupamiento KMEANS y CLARA, para ellos todos los datos empleados fueron previamente discretizados convirtiéndolos en datos numéricos. En ambos casos el procedimiento a seguir estuvo formado por dos pasos. Primero se aplicó la técnica de agrupamiento considerando las características de las tareas.

Aquellos grupos con pocas tareas fueron analizados porque pueden representar tareas muy específicas o tareas anómalas. En un segundo paso se determinan en cada grupo aquellas tareas que tienen diferencias significativas respecto al costo con otras tareas de su mismo grupo las cuales son identificadas como tareas anómalas, este último análisis se realizó basado en el análisis univariado del atributo costo. Como criterio de comparación se determinó la cantidad de datos anómalos identificados.

Tabla 3: Comparación de método tradicional con el método propuesto para análisis riesgos.

	Método PMBok				Método propuesto										
	Experto				Experto1			Experto2			Experto3			Exp	
	P	I	Exp	Ord	P	I	D	P	I	D	P	I	D		
Pérdida de recursos humanos, condiciones	0.5	0.6	0.3	6	Alta	Alta	Muy baja	Media	Alta	Baja	Perfecta	Alta	Ninguna	Media, 0.11	
Bajo de nivel de formación de RH	0.3	0.6	0.18	9	Media	Alta	Baja	Media	Alta	Ninguna	Muy alta	Alta	Ninguna	Media, -0.22	
Pocos incentivos al equipo, baja producción	0.8	0.6	0.48	1	Alta	Media	Muy baja	Alta	Alta	Ninguna	Media	Muy alta	Muy baja	Media, -0.22	
Mala conformación de equipos	0.4	0.6	0.24	8	Media	Media	Media	Media	Baja	Baja	Muy baja	Media	Media	Media, -0.44	
Cliente desinteresado, que no participa en los encuentros.	0.2	0.8	0.16	10	Baja	Alta	Muy baja	Baja	Alta	Ninguna	Baja	Perfecta	Baja	Med, -0.44	
Tardía entrega de información por cliente	0.3	0.9	0.27	7	Baja	Muy alta	Baja	Baja	Alta	Ninguna	Media	Alta	Muy baja	Media, -0.44	
Atrasos en entrega de los proveedores	0.4	0.9	0.36	4	Alta	Muy alta	Media	Baja	Alta	Ninguna	Alta	Alta	Media	Media, 0.22	
Aumento de los precios de los recursos	0.3	0.3	0.09	13	Baja	Alta	Alta	Muy baja	Alta	Med	Alta	Muy alta	Alta	Media, 0.44	
Dificultades energéticas, afectan la producción	0.6	0.6	0.36	3	Alta	Muy alta	Baja	Media	Muy alta	Ninguna	Muy alta	Muy alta	Muy alta	Alta, -0.22	
Dificultades con el transporte, afecta plan	0.5	0.7	0.35	5	Med.	Muy alta	Baja	Media	Muy alta	Perfecta	Alta	Perfecta	Baja	Alta, 0	
Fenómenos atmosféricos afectan la productividad	0.1	0.9	0.09	12	Muy baja	Perfecta	Muy alta	Muy baja	Perfecta	Muy baja	Baja	Muy alta	Media	Media, -0.33	
Rotura de equipos y lento mantenimiento.	0.6	0.7	0.42	2	Media	Muy alta	Alta	Med.	Alta	Alta	Muy alta	Alta	Muy baja	Alta, -0.33	
Trámites engorrosos para la comercialización	0.5	0.2	0.1	11	Media	Alta	Baja	Media	Media	Media	Alta	Perfecta	Media	Med, 0.44	

En la comparación de los resultados se usó el tests de Wilcoxon y se compararon los algoritmos CLARA y KMeans, se encontraron diferencias significativas entre las muestras. Aportando mejores resultados el algoritmo basado en CLARA, ver Tabla 4.

Tabla 4. Muestra de código R en la comparación de los resultados finales, de los dos algoritmos.

```
result_outliers_tasts_info = summary(data_all)
>result_outliers_tasts_info
Corrida Kmeans.Correctos KMEANS CLARA.Correctos CLARA
Min. : 1.00 Min. :238.0 Min. :63.98 Min. :342.0 Min. :91.94
1st Qu.: 3.75 1st Qu.:267.2 1st Qu.:71.84 1st Qu.:346.8 1st Qu.:93.21
Median : 6.50 Median :285.0 Median :76.61 Median :351.0 Median :94.36
Mean : 6.50 Mean :293.0 Mean :78.76 Mean :353.2 Mean :94.96
3rd Qu.: 9.25 3rd Qu.:310.2 3rd Qu.:83.40 3rd Qu.:361.0 3rd Qu.:97.04
Max. :12.00 Max. :357.0 Max. :95.97 Max. :370.0 Max. :99.46
>with( data_all, wilcox.test(KMEANS, CLARA, paired=TRUE))
Wilcoxon signed rank test
data: KMEANS and CLARA
V = 4, p-value = 0.003418
alternative hypothesis: true location shift is not equal to 0
```

A partir del análisis de los resultados, se recomienda que en los análisis se experimente empleando técnicas de detección de datos anómalos colectivos. Porque en la práctica algunos proyectos podrían tener costos sobregirados en todas las tareas provocando un enmascaramiento en las tareas dentro de un mismo proyecto pero que podrían ser detectadas como anómalas en caso de que se analicen por proyecto. Se deben experimentar con técnicas basadas en el ensamblaje de diferentes métodos para potenciar la detección de datos anómalos en diversas situaciones.

Conclusiones

Se identificó como tendencia en las estrategias de aseguramiento de ingresos la necesidad de combinar las estrategias reactivas, con estrategias activas y proactivas, con el objetivo de disminuir los tiempos de detección de las fugas de ingresos y prevenir las posibles fallas o acciones de fraudes. Además, se identificó que en la implantación de procesos de aseguramiento de ingresos es preciso el desarrollo de actividades de formación en los principales estándares, en las tecnologías y en los sistemas de información que registran los datos de los procesos organizacionales.

Se considera que los métodos para el aseguramiento de ingresos deben ser ajustados a las organizaciones donde se aplican. De esta misma forma, los métodos de detección de datos anómalos como parte de las estrategias de aseguramiento de ingresos, necesitan ser orientados al problema específico que se quiere resolver.

Se realizaron experimentos asociados a comprobar los algoritmos empleados en diferentes componentes del método propuesto. En particular en el caso de la evaluación cualitativa de los riesgos el método propuesto basado en computación con palabras reportó mejores resultados que el método tradicional publicado en el estándar PMBok. Para el caso de las comparaciones realizadas en la detección de datos anómalos se encontraron mejores resultados con el uso del agrupamiento CLARA que con los resultados obtenidos con KMeans.

Se recomienda que en los análisis se experimente empleando técnicas de detección de datos anómalos colectivos y con técnicas basadas en el ensamblaje de diferentes métodos para potenciar la detección de datos anómalos en diversas situaciones.

Referencias

- ACOSTA, K. Aseguramiento de ingresos: una actividad fundamental en las empresas de telecomunicaciones. Revista Industrial, 2008, 29(2).
- AGGARWAL, CH. Outlier Analysis. Springer, New York, USA, ISBN 978-1-4614-6395-5, DOI 10.1007/978-1-4614-6396-2, 2013.
- CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly Detection for Discrete Sequences: A Survey. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(5): p. 823–839.
- CNT, Portal colaborativo, Departamento de Aseguramiento de Ingresos. 2015. Consultado en: junio de 2016. Disponible en: <http://corporativo.cnt.gob.ec/cnt-ep-contribuye-con-el-estado/>.
- DENESHKUMAR, V.; SENTHAMARAIKANNAN, K.; Manikandan, M. Identification of Outliers in Medical Diagnostic System Using Data Mining Techniques. International Journal of Statistics and Applications, DOI: 10.5923/j.statistics.20140406.01, 2014, 4(6). Disponible en: <http://www.researchgate.net/publication/274721695>.
- GOGOI, P., et. al. Outlier Identification Using Symmetric Neighborhoods. 2nd International Conference on Communication Computing & Security [ICCCS], doi: 10.1016/j.protcy.2012.10.029, 2012, 6, p. 239 – 246.
- GUPTA, M.; GAO, J., et. al. Outlier Detection for Temporal Data. ISBN(paper): 9781627053754, ISBN(ebook): 9781627053761, 2014. Disponible en: www.morganclaypool.com.

GILBERTO, C. A. Un modelo para el aseguramiento de ingresos basado en técnicas de soft computing. Trabajo para optar por el grado de Doctor en Ciencias Técnicas. Ciudad de La Habana, septiembre de 2015.

KARANJIT, S.; SHUCHITA, U. Outlier Detection: Applications And Techniques.[En línea]. IJCSI International Journal of Computer Science Issues, ISSN: 1694-0814, 2012, 9(3). Disponible en: <http://www.IJCSI.org>.

KHAN, N. Internship Report on Revenue Assurance and Fraud Management. ID: 10104009. 2014.

KUNA, H. D., et al. Outlier detection in audit logs for application systems. Information Systems, DOI: 10.1016/j.is.2014.03.1, 2014, 44, p. 22–33. Disponible en: <http://www.researchgate.net/publication/262915159>.

LI, Z.; ROBERT, J., et. al. A Unified Framework for Outliers Detection in Trace Data Analysis. IEEE Transactions on semiconductor manufacturing, DOI: 10.1109/TSM.2013.2267937, 2014, 27(1).

MASSYN, R. H. A provisional taxonomy of revenue assurance: a grounded theory approach, M.Sc. in Philosophy, University of Johannesburg, TH 621.3820683, 2010. Disponible en: <https://ujdigispace.uj.ac.za>.

MATTISON, R. The Revenue Assurance Standards Release 2009. Published by XiT Press, Oakwood Hills, Illinois, USA, Copyright Rob Mattison. 2009.

PIÑERO, P.; TORRES, S., et al. GESPRO. Paquete para la gestión de proyectos. Revista Nueva Empresa, Cuba, ISSN: 1682-2455, 2013, 9(1): p. 45-53.

PIVIDORI, M.; STEGMAYER, G.; Milone, D. Cluster Ensembles for Big Data Mining Problems. 2015. Disponible en: <http://www.researchgate.net/publication/281461828>.

SATHISHKUMAR, E.; THANGAVEL, K. A Novel Approach for Outlier Detection using Rough Entropy. Department of Computer Science Periyar University, In WSEAS TRANSACTIONS on COMPUTERS, E-ISSN: 2224-2872, 2015, 14.

SHIVANI, P.; SHAH, V.; VALA, J. Outlier Detection in Dataset using Hybrid Approach. International Journal of Computer Applications (0975 – 8887), 2015, 122(8).

SHPIGELMAN. A Unified Framework for Outlier Detection in Trace Data Analysis. IEEE Transactions on semiconductor manufacturing, DOI: 10.1109/TSM.2013.2267937, 2014, 27(1).

THE STANDISH GROUP INTERNATIONAL. The CHAOS Manifesto. The Standish Group International, Incorporated. 2013.

THE STANDISH GROUP INTERNATIONAL. Chaos Report. New York: The Standish Group International, Inc. 2014.

TM FORUM. Revenue Assurance practitioner blog: Do we need a new approach to revenue assurance in the digital world? & seeing is believing: Setting revenue assurance KPIs. 2015.

VIJENDRA, S.; PATHAK S. Robust Outlier Detection Technique in Data Mining: A Univariate Approach. Faculty of Engineering and Technology, Mody Institute of Technology and Science, Lakshmangarh, Sikar, Rajasthan, India 2013.

ZIMEK, A., et. al. A Survey on Unsupervised Outlier Detection in High-dimensional Numerical Data. Journal on Statistical Analysis and Data Mining, 2012. Disponible en: <http://onlinelibrary.wiley.com/doi/10.1002/sam.11161/abstract>.

ZHANG, J.; ZHANG, S., et al. An outlier mining algorithm based on constrained concept lattice. International Journal of Systems Science, DOI: 10.1080/00207721.2012.745029, 2014, 45(5): p. 1170-1179. Disponible en: <http://www.researchgate.net/publication/262317735>.