

Identificación de factores que permitan potencializar el éxito de proyectos de desarrollo de software

Identification of factors that allow potentialize the success of development projects software.

Mauricio Gallego Gallego¹, Javier Hernández Cáceres²

¹Ingeniero de sistemas y computación, Universidad Autónoma de Manizales, Manizales, Colombia
ingmauricio05@gmail.com

²Departamento de Matemáticas, Universidad Santo Tomas, Bucaramanga, Colombia
javiermineria@gmail.com

Resumen— Lograr el éxito en los proyectos de desarrollo de software es una tarea a la que día a día se están enfrentando las empresas productoras de software, estudios realizados por diferentes organizaciones demuestran que la tasa de éxito de los proyectos esta alrededor del 32%, y de proyectos fallidos alrededor del 24% [1], siendo la diferencia de estos una cifra relativamente baja lo que conlleva a que se deba evaluar el tema y se piense en tomar medidas; el éxito de los proyectos depende de la sincronización que se pueda tener entre diferentes factores como los humanos, los técnicos, los económicos, tiempo, alcance, entre otros. En este artículo se realiza un análisis de factores de proyectos exitosos y fallidos de una empresa de outsourcing de TI aplicando técnicas de minería de datos, con el fin de obtener patrones que permitan a la organización tomar decisiones y guiar los proyectos de software a la consecución del éxito.

Para lograr tal objetivo la investigación se desarrolló siguiendo la metodología CRISP-DM, que es una metodología utilizada como referente para los proyectos de minería de datos. Finalmente se logró obtener un modelo de regresión logística que permitió determinar que los factores que inciden en el éxito de un proyecto están expresados por:

Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto, y el tamaño del equipo de proyecto; siendo estos los 2 únicos factores de los 27 estudiados capaces de explicar el éxito o fracaso de los proyectos.

Palabras clave— CRISP-DM, Factores de éxito, Minería de datos, Proyectos de software exitosos, Regresión logística.

Abstract— Achieving success in software development projects is a task that day are facing companies producing software , studies by various organizations show that the success rate of projects is about 32% , and failed projects around 24% [1], the difference of these relatively low figure which entails that the subject should evaluate and consider taking action; the success of the project depends on the synchronization can be between different factors as human , technical , financial , time , scope, among others. This article analyzes factors of successful and failed projects of a company outsourcing applying data mining techniques in order to obtain patterns that enable the

organization to make decisions and lead software projects to achieve success.

To achieve this goal the research was conducted following the CRISP- DM methodology , a methodology used as a reference for data mining projects . Finally managed to get a logistic regression model that allowed us to determine the factors that influence the success of a project are expressed by:

Excellent definition of roles and responsibilities for project development activities, and the size of the project team ; these being the only 2 of the 27 studied factors able to explain the success or failure of projects.

Key Word —CRISP-DM, Success Factors, Data Mining, successful software projects, Logistic Regression.

I. INTRODUCCIÓN

Con el paso de los años y la evolución tecnológica, la innovación y la facilidad de agilizar los procesos operativos diarios del ser humano; el mercado mundial ha convergido en gran porcentaje al desarrollo de proyectos de software. Gran cantidad de empresas internacionalmente han extendido su nombre gracias al éxito de sus sistemas informáticos que soportan el *core* de importantes compañías mundiales. Pero a pesar de que se encuentran en el tope del éxito, no todas sus iniciativas de software han alcanzado y obtenido buenos resultados; de allí que incluso en esta era donde abundan los marcos de trabajo, estudios, procesos y metodologías entre otros para garantizar la consecución plena de sus objetivos permanece la duda de cómo garantizar el éxito de un proyecto. El estudio del éxito de los proyectos de software ha sido un tema de gran atención desde hace más de una década, en la que se ha tratado de analizar los factores que colaboran al cumplimiento completo de un desarrollo y la manera como las empresas tratan de evitar posibles fracasos. Un ejemplo fue el análisis realizado [1] sobre las estadísticas del Standish Group de proyectos exitosos donde evidenciaba que el crecimiento

de efectividad de los proyectos entre 2000 y 2004 había sido solo de un 1%, situación que mejora al 2009 con un 4%. Cabe aclarar que con el paso de los años los factores tenidos en cuenta para el estudio han estado ampliándose para hallar con más efectividad una respuesta [2]. De 1999 donde se consideraban solo 5 elementos, para el 2012 han sido duplicados.

Un análisis desde otro flanco, incluyendo la cara oscura en el desarrollo de proyectos de software, la variación respecto a fracasos o rescate de los mismos si registra grandes diferencias. Como referencia entre los años 1994 a 2008, los proyectos fallidos pasaron de un 31% a 24%, y proyectos medianamente cumplidos registraron una diferencia de 7% menos respecto a 1994 donde equivalían al 31% [3].

Los resultados son alentadores, sin embargo no van en tendencia creciente, registrando años en donde el éxito, fracaso o mediano cumplimiento aumenta o disminuye variablemente. Esto a qué se debe?, si bien, aunque cada entidad consultora ha ido perfeccionando la respuesta a los factores de consecución de los proyectos de software, todas convergen en tres principios principales: tiempo, costo y calidad [4].

Es claramente entendido que un proyecto de software es exitoso cuando el presupuesto estimado no es sobrepasado, el tiempo de entrega propuesto se cumple con exactitud y han cumplido con los objetivos propuestos; sin embargo es necesario dar atención a demás elementos que hacen parte para que estos componentes se alcancen.

Un proyecto de software está conformado básicamente por desarrolladores, analistas de pruebas, analistas de requerimientos y jefes de proyecto, de quienes se espera un buen entendimiento, trabajo en equipo, disponibilidad, liderazgo, motivación, experiencia y skills aptos para afrontar los retos. Podría considerarse entonces que si una de estas cualidades o características está ausente o falla dentro del grupo de trabajo, alguno de los tres principios de éxito perderá los lineamientos. Fallas como: pobre estructura organizacional, falta de liderazgo, falta de apoyo del nivel gerencial, falta de esfuerzo, choques de personalidades; uso inefectivo de métodos de desarrollo de software, procesos de negocios y asignación de recursos inapropiados, gestión de proyectos y herramientas de seguimiento inadecuados [5].

Además de todas las propuestas que se encuentra en la teoría, es importante contar con herramientas que permitan obtener ideas simples de lo que fundamentalmente se requiere para entablar buenos cimientos que colaboren a garantizar un proyecto exitoso; de aquí el poder basarse en juicios de expertos o iniciativas anteriores que hayan cumplido con las necesidades, tanto para el proveedor como el cliente.

Dentro de empresas de gran magnitud, cuya fuente de sostenibilidad es el desarrollo de sistemas de información para clientes de altas exigencias, aprovechar el conocimiento de experiencias pasadas es un punto a favor para procurar no repetir errores y mitigar cada vez más un posible fracaso [6].

En el presente trabajo, se abordará desde la gestión del conocimiento [6] técnicas para la obtención de información y datos dentro de la empresa consultora, que permitan

determinar factores de éxito específicos para los diferentes proyectos de software, bien sea por la cantidad de personal requerido, las habilidades necesarias para afrontar los proyectos, la experiencia en las herramientas requeridas para el desarrollo de los requisitos, cualidades de los trabajadores entre otros.

Por medio de la minería de datos se busca conocer las razones y consecuencias de los proyectos principales de la compañía que hayan sido críticos en su progreso y crecimiento; pero sin dejar atrás aquellos que hayan sido un punto de inflexión donde sus metas hayan sido obstruidas.

A partir de la información obtenida, se podrá entonces identificar patrones que permitan ubicar las iniciativas de desarrollo en diferentes ambientes con el fin de procurar buenos resultados de los mismos. Se realizará un análisis según las diversas características que estos presenten tomando como referencia la historia y experiencias de los casos de éxito y fracaso de la empresa.

Todo este trabajo se realizará según los lineamientos expuestos por la metodología CRISP-DM; finalmente se presentan los resultados, análisis, y conclusiones.

II. CONTENIDO

Se define la minería de datos como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde almacenes de datos en distintos formatos, con el fin de generar patrones que permitan la toma de decisiones en la organización y por lo tanto generar beneficios a la misma; en este sentido la minería de datos se enfrenta a un gran reto, el uso de técnicas adecuadas que permita analizar estos datos y extraer conocimiento útil y novedoso [7].

A. Metodología

Existen varias metodologías para orientar el proceso de minería de datos. Entre esas metodologías se destacan CRISP-DM y SEMMA (Sample, Explore, Modify, Model, Asses). SEMMA se centra en las características técnicas del desarrollo del proceso, mientras que CRISP-DM mantiene como foco central los objetivos empresariales del proyecto. Debido a ello, CRISP-DM comienza realizando un análisis del problema empresarial para su transformación en un problema técnico de minería de datos. CRISP-DM define una estructura para proyectos de minería de datos y suministra la orientación para su ejecución [8].

Para este proyecto de investigación aplicada, la metodología seleccionada fue CRISP-DM, por el éxito presentado en los proyectos de minería de datos y porque que abarca todos los objetivos propuestos.

El ciclo de vida del proyecto según la metodología CRISP-DM está basado en 6 fases cambiantes entre sí, lo cual lo postula como un ciclo en constante movimiento [9].

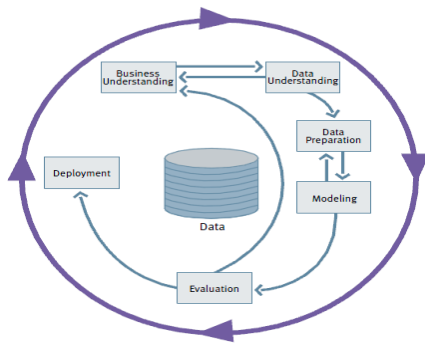


Fig. 1: Metodología CRISP-DM

Los datos fueron recolectados partiendo de la identificación los factores que influyen en la inyección de defectos de desarrollo de software, ya que había un interés de la compañía de evaluar cuales de estos factores podían aplicar en los proyectos de la compañía e identificar y analizar cuales influían en la consecución de proyectos de software exitosos; se realizó una encuesta vía internet de pregunta cerrada a los involucrados en el desarrollo de los proyectos de software basada en los factores identificados en la primera sección.

La población objetivo hacia la cual estaba orientada la encuesta era a los líderes técnicos o arquitectos de cada equipo de trabajo, se seleccionó esta persona ya que se considero que es la persona que tiene el conocimiento tanto técnico como de gestión de los proyectos que han culminado, adicionalmente, aunque hay mucha rotación de personal en la compañía, el arquitecto es uno de los cargos que se conservan con más tiempo y por lo tanto da la opción de que se tenga mayor cantidad de información de los proyectos terminados y así poder brindar una información sustancial y valedera, ya que la primicia para un proyecto de minería de datos parte de la cantidad de datos y consistencia de información que se tenga; la idea finalmente era poder evaluar todos los proyectos que hayan sido culminados y de los cuales se pudo obtener la información requerida por medio de encuestas.

El total de proyectos culminados de los cuales se tuvo información para evaluar por medio de las encuestas fue de 83.

1. *Compresión del negocio.*

Objetivos institucionales

- ✓ Descubrir factores de éxito de los proyectos de la compañía.
- ✓ Mejorar el rendimiento de los proyectos de software.
- ✓ Obtener ejecuciones futuras de proyectos exitosos.
- ✓ Disminuir la cantidad de ejecución de proyectos no exitosos.

Evaluación de la situación

- ✓ Recursos computacionales: La realización de la encuesta se hizo por medio de la herramienta form de

Google Drive, la cual brindó facilidad para ejecutar la encuesta y suplió las necesidades de la misma.

Para el análisis y procesamiento de la información de las encuestas, se contó con un equipo portátil Dell Studio 1557, con procesador intel core i7 de 1.60GHz y 3Gb de memoria ram; adicionalmente, el uso de software y herramientas gratuitas o versiones trial de las mismas.

- ✓ Datos: después de realizar el estudio bibliográfico [10][11][12] de los factores que inciden en la inyección de defectos en el desarrollo de software, y posterior análisis con el líder de arquitectura y procesos de la compañía se seleccionaron los siguientes factores, sobre los cuales él considera serían de mayor interés y viables a evaluar en las encuestas, en algunos casos algunos factores identificados en el estudio bibliográfico se dividieron o renombraron, con el fin de facilitarle el entendimiento por parte de los encuestados y de ampliar la cobertura del factor:

- Conocimiento de los desarrolladores acerca de los procesos de desarrollo.
- Grado de experiencia con la plataforma / entorno del proyecto.
- Grado de experiencia con herramientas y lenguaje de desarrollo.
- Conocimientos de los desarrolladores sobre las herramientas usadas para soportar actividades de desarrollo del proyecto.
- Alta capacidad de los desarrolladores para entender la arquitectura y resolver los riesgos asociados del proyecto.
- Experiencia previa de los desarrolladores con aplicaciones similares al proyecto.
- Conocimientos de los desarrolladores sobre los aspectos de negocio del producto a desarrollar.
- Conocimientos de los desarrolladores sobre el dominio para el que está destinado el producto.
- Conocimientos de los desarrolladores sobre el producto a desarrollar.
- Capacidad de comunicación: Se dividió en dos:
 - Capacidad de comunicación de los desarrolladores.
 - Capacidad de comunicación de los diseñadores
- Capacidad de los desarrolladores: se dividió en dos:
 - Capacidad de los desarrolladores para resolver problemas.
 - Capacidad de los diseñadores para resolver problemas.
- Nivel de habilidad de los desarrolladores: se renombró por:

- Conocimientos técnicos de los desarrolladores.
- Nivel de habilidad de los diseñadores: se renombró por:
 - Conocimientos técnicos de los diseñadores.
- Complejidad de las interfaces externas del producto a desarrollar.
- Definición de roles y responsabilidades para las actividades de desarrollo del proyecto.
- Definición de responsabilidades y niveles de autoridad para las actividades de desarrollo del proyecto.
- Adherencia a los planes: se dividió en 2:
 - Cumplimiento con el plan de trabajo respecto al proyecto.
 - Cumplimiento con el plan de trabajo respecto a los procesos de la empresa.
- Experiencia en programación de los desarrolladores.
- Experiencia del equipo en proyectos de desarrollo.
- Uso de personas con postgrado en el proyecto.
- Tamaño del equipo del proyecto.
- Nivel de continuidad y estabilidad del equipo del proyecto.
- Cantidad de interfaces externas del producto (tanto de hardware como software).
- Tamaño del producto a desarrollar.

- ✓ Supuestos: Los líderes técnicos o arquitectos fueron las personas seleccionadas para responder las encuestas, ya que se consideró que tenían todo el conocimiento de la ejecución de los proyectos, tanto en la parte técnica como en la parte de gestión, y en base a esto podrán responder la encuesta de una manera más acertada.

Objetivos de la minería de datos

- ✓ Identificar y determinar los factores que influyen en la consecución de proyectos de software exitos.

2. Compresión de los datos.

Recolección inicial de datos

Se ejecutaron 83 encuestas para identificar el comportamiento que han tenido los 30 factores seleccionados en la ejecución de los proyectos.

- ✓ Tipo: Encuesta de pregunta cerrada usando escala de *Likert* de 1 a 5 para medir el comportamiento de los factores en los diversos proyectos.
- ✓ Locación: Encuesta online, usando la herramienta *Form* de la aplicación *Google Drive*.
- ✓ Métodos de acceso: Internet.

Descripción de datos

La calificación de los factores en las encuestas estaba dada por una escala de *Likert* de 1 a 5, donde:

- 1: Totalmente en desacuerdo
- 2: Parcialmente en desacuerdo
- 3: Ni de acuerdo ni en desacuerdo.
- 4: Parcialmente de acuerdo
- 5: Totalmente de acuerdo

Verificar la calidad de los datos

Con el fin de evitar el sesgo en las encuestas, la pérdida, errores y falta de datos, se hicieron las encuestas de manera online controlando la obligatoriedad de las preguntas en todo el cuestionario, así mismo, solo se tendrían datos en un rango de 1 a 5, lo que permitió tener controlado el manejo de la información. De igual forma se revisó que todos los datos resultados tuvieran valores comprendidos entre 1 y 5, y que no hubiera datos perdidos.

3. Preparación de los datos.

Selección de datos y limpieza de datos

Se excluyeron del análisis y modelamiento de los datos las variables conformadas por los factores: Ejecución del proyecto dentro del tiempo esperado, Ejecución del proyecto dentro del presupuesto esperado, Ejecución del proyecto dentro del alcance esperado, debido a que la información de estos factores se recolectó solo con el único fin de recrear la variable ProyectoExitoso, adicionalmente no hacen parte de los factores que inciden en la inyección de defectos de desarrollo de software.

Construcción de datos

Se construyó la variable proyecto exitoso, como el promedio de las variables Ejecución del proyecto dentro del tiempo esperado, ejecución del proyecto dentro del presupuesto esperado y ejecución del proyecto dentro del alcance esperado, se tomaron en cuenta solo estas variables ya que la definición de éxito para esta investigación estaba contenida dentro del tiempo, costo y alcance.

Después de obtener el promedio, como el fin de la investigación era realizar una predicción de una variable dependiente partiendo de n variables dependientes, donde se debe identificar si es exitoso o no, se determinó que si el

promedio resultado era igual o superior a 4 el proyecto se calificaba con 1, que representaba exitoso, o si el resultado era inferior a 4 se calificaba con 0, que representaba fracaso.

Integración de datos

Como solo se tenía una fuente de datos, que era el archivo Excel con el resultado de las encuestas, no se requirió hacer una combinación de datos.

4. Modelado.

Seleccionar la técnica de modelado

Se selecciono el modelo de regresión logística binomial, se hizo teniendo en cuenta la naturaleza de la variable dependiente ya que era categórica binaria, es decir, que solo puede tomar dos valores, que para nuestro caso fue las variable proyecto exitoso cuyos valores fueron (0) para proyectos no exitosos y (1) para el proyectos exitosos. Así mismo las variables independientes podian ser de cualquier naturaleza ya sea cualitativa o cuantitativa.

Por otro lado la idea era construir un modelo y obtener una ecuación con fines de predicción, partiendo de la selección de n variables independientes.

Generar el diseño de las pruebas

La pruebas ómnibus valida si la varianza explicada en un conjunto de datos es significativamente mayor que la varianza no explicada, en general. Se asume que es estadísticamente significativa si el valor es ≤ 0.05 .

La prueba de bondad de ajuste va a sugerir en el modelo no existen diferencias significativas entre los valores observados y los pronosticados.

La prueba de Bartlett, la cual está referida a la matriz de correlaciones. Se contrasta la siguiente hipótesis nula (Ho): La matriz de correlaciones es una matriz de identidad; versus la hipótesis alternante: la matriz de correlaciones no es una matriz de identidad. En caso de rechazar la Ho se concluye que las variables están correlacionadas entre sí, lo que da sentido al análisis componentes principales a realizar.

La prueba de Kaiser-Meyer-Olkin (KMO) estima un valor que de acuerdo a su ubicación en una escala permitirá concluir si el análisis realizado es conveniente. En la medida que los primeros sean más altos, el valor estimado estarán más cerca de uno, y por tanto el modelo factorial empleado será más efectivo.

Para interpretar la prueba de Bartlett se tiene en cuenta qué:

- 1 \geq KMO > 0.9 excelente
- 0.9 \geq KMO > 0.8 bueno
- 0.8 \geq KMO > 0.7 aceptable
- 0.7 \geq KMO > 0.6 regular
- 0.6 \geq KMO > 0.5 deficiente
- KMO \leq 0.5 inaceptable

Se utilizó el método de análisis de spearman para el estudio de las correlaciones debido a la naturaleza de las variables involucradas, las cuales son de naturaleza cualitativa y cuyo nivel de medición es ordinal, y la idea era identificar los factores menos correlaciones entre sí. La correlación es una medida de asociación (fuerza) de la relación entre dos variables. Que varía de 0 a 1 (relación lineal perfecta) o -1 (perfecta relación lineal negativa).

Construcción del modelo

Para la construcción de los modelos de regresión logística se utilizó el 95% de los datos obtenidos lo que represento trabajar con 77 registros, los demás 6 registros fueron utilizados para probar el modelo; los factores seleccionados para la construcción del modelo fueron: Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto y Tamaño del equipo del proyecto.

Ahora se presentan las funciones del análisis de regresión logística en la tabla 1. Se puede observar en ella que las variables incluidas en cada modelo son significativas; es decir, cada coeficiente (B) es diferente a cero, lo cual sugiere que dicha variable tiene un valor en cada función discriminante de regresión logística. Otro aspecto que corrobora lo anterior es el hecho de que el intervalo de confianza del 95% para $Exp(B)$, no es 1. Así mismo el signo de cada coeficiente (B) nos está sugiriendo la dirección en la cual dicha variable discrimina a los proyectos; por ejemplo en el modelo 2, el coeficiente (B) de 1.064 para la variable Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto nos indica que un aumento en una (1) unidad en esta variable, aumenta en un 1.064 la probabilidad de que un proyecto pertenezca al grupo de exitoso; y lo contrario sucede para el coeficiente (B) de -.713 para la variable Tamaño del equipo del proyecto, la cual sugiere que un aumento en una (1) unidad en esta variable, disminuye en un 0.713 la probabilidad de que un proyecto pertenezca a este mismo grupo. Dicho en otras palabras, un proyecto exitoso tiende a ser diferenciado de uno no exitoso por tener de manera simultánea mayor Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto y Menor Tamaño del equipo del proyecto.

Representación del modelo de regresión logística bivariado

Partiendo de lo que se pretende mediante la Regresión Logística es expresar la probabilidad de que ocurra el evento en cuestión como función de ciertas variables, que se presumen relevantes o influyentes. Si ese hecho que queremos predecir lo representamos por Y (la variable dependiente), y las k variables explicativas (independientes y de control) se designan por X1, X2, X3,...,Xk, la ecuación general (o función logística) es:

$$P(Y=1) = \frac{1}{1 + \exp(-\alpha - \beta_1X_1 - \beta_2X_2 - \beta_3X_3 - \dots - \beta_KX_k)}$$

Donde α , β_1 , β_2 , β_3, \dots, β_k son los parámetros del modelo, y \exp denota la función exponencial. Esta función exponencial es una expresión simplificada que corresponde a elevar el número e a la potencia contenida dentro del paréntesis, siendo

e el número o constante de Euler, o base de los logaritmos neperianos (cuyo valor aproximado a la milésima es 2,718).

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
								Inferior	Superior
Paso 1 ^a	Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto	.989	.260	14.468	1	.000	2.688	1.615	4.474
	Constante	-3.396	1.019	11.098	1	.001	.033		
Paso 2 ^b	Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto	1.064	.293	13.209	1	.000	2.898	1.633	5.143
	Tamaño del equipo del proyecto	-.713	.247	8.342	1	.004	.490	.302	.795
	Constante	-.997	1.306	.583	1	.445	.369		

Tabla 1: Variables en la ecuación del modelo

Ahora bien, para nuestro modelo la función logística fue de la siguiente manera:

$$P(Y=1) = \frac{1}{1 + 2.718^{(-(-0.997) - (1.064(X1)) - (-0.713(X2)))}}$$

Donde X1= Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto.

X2 = Tamaño del equipo del proyecto

Evaluación del modelo

La técnica utilizada para construir el modelo fue de paso a paso hacia delante por ser el de mayor precisión, mediante el procedimiento de máxima verosimilitud.

Para evaluar de forma global la validez modelo tenemos los indicadores de la tabla 2: el cual sugiere que en el 1 primer paso tenemos la inclusión del primer factor (Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto) y el índice de R cuadrado de Nagelkerke, sugiere que dicha variable explica el 27.8% de la variación de la variable dependiente (calificación del proyecto), mientras en el segundo paso con la inclusión del segundo factor (Tamaño del equipo del proyecto) el modelo mejora ostensiblemente al alcanzar un 40.7% de la variación de la variable dependiente la cual es explicada por los factores incluidos en el modelo.

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	86,049 ^a	,206	,278
2	76,197 ^b	,301	,407

Tabla 2: Resumen del modelo

Los resultados de esta prueba arrojaron dos modelos de inclusión de variables: en el primer modelo se incluyó únicamente la variable Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto; en el segundo modelo se incluyó la variable, Tamaño del equipo del proyecto. Es de recordar que cada vez

que se incluye una variable nueva en cada modelo, las demás permanecen en él; es decir en el modelo dos quedaron las variables Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto y Tamaño del equipo del proyecto. Es importante señalar, que aunque el factor Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto presentó fue seleccionada para ser parte de este análisis, aquí en el análisis multivariado se excluyó por no aportar suficiente información para la construcción de los modelos, puntuación de 6,369 para una sig. 0.12

		Puntuación	gl	Sig.
Paso 0	Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto	16,930	1	,000
	Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto	6,369	1	,012
	Tamaño del equipo del proyecto	10,016	1	,002
	Estadísticos globales	25,372	3	,000

Tabla 3: Variables que no están en la ecuación

En la Tabla 4 sobre los coeficientes del modelo podemos señalar que los dos modelos son estadísticamente significativos al tener un valor ≤ 0.02 ; es decir, las variables independientes tomadas juntas tienen un efecto discriminante altamente significativo sobre la variable dependiente. Es decir ayudan a predecir cual pertenece a éxito y cual a fracaso.

		Chi cuadrado	gl	Sig.
Paso 1	Paso	17,755	1	,000
	Bloque	17,755	1	,000
	Modelo	17,755	1	,000

Paso 2	Paso	9,852	1	,002
	Bloque	27,607	2	,000
	Modelo	27,607	2	,000

Tabla 4: Pruebas ómnibus sobre los coeficientes del modelo

De igual forma la prueba de bondad de ajuste (tabla 5) sugiere que en cada modelo no existen diferencias significativas entre los valores observados y los pronosticados; lo que implica que en cada modelo estimado los datos observados y los datos esperados son similares en buena medida. Es de recordar, en este aspecto, que en la medida en que el valor de la significancia se encuentra mucho más alto de 0.05, el ajuste de los datos es mucho mejor; y en este caso el primer modelo fue el que más se ajustó a los datos.

Paso	Chi cuadrado	gl	Sig.
1	,516	2	,773
2	5,679	7	,578

Tabla 5: Prueba de Hosmer y Lemeshow

Para la evaluación del modelo se utilizó el 5% de los datos (6 proyectos) dado que para la construcción del modelo se utilizó el 95% de los datos restantes (77 proyectos). Aplicando la ecuación que describe el modelo, se obtuvo un valor que va de 0 a 1, para determinar si un proyecto es exitoso, este valor como mínimo deberá superar un valor de 0.6, mientras un proyecto no éxitos estará por debajo de dicha cifra. Entre más cercana 1 sea el valor calculado más probabilidad de que el proyecto sea exitoso, por el contrario entre más cercano a 0 más probabilidad de que el proyecto no sea éxitos. Los resultados de realizar lo anterior se muestran en la tabla 6 en ella se puede apreciar que el modelo pronostico acertó en la clasificación de proyecto exitoso en 1 de 2 dos, mientras en los proyectos no éxitos logro clasificar 3 de 4, en general el modelo logro clasificar correctamente 4 proyectos de 6, es decir, tuvo un 66.6% de aciertos.

		Calificación Proyecto		Total
		Fracaso	Exitoso	
Pronostico	Fracaso	3	1	4
	Exitoso	1	1	2
Total		4	2	6

Tabla 6: Contingencia 1

Simulación

Se realizó una simulación del modelo aplicándolo a toda la base de datos (resultados de las encuestas), se encontró que el modelo pronostico logro clasificar correctamente 37 proyectos éxitos de un total de 48, es decir, logro identificar correctamente el 77% de los proyectos éxitos, por otra parte, logro identificar 25/35, es decir, un 71% de los proyectos no éxitos, en general, el modelo pudo identificar correctamente 62 proyectos de 83 posibles, en otras palabras

el modelo pudo identificar correctamente el 74% de los proyectos evaluados.

		Calificación Proyecto		Total
		Fracaso	Exitoso	
Pronostico	Fracaso	25	11	36
	Exitoso	10	37	47
Total		35	48	83

Tabla 7: Simulación

5. Evaluación.

Análisis multivariado

A continuación se presentan los factores representativos de cada categoría que fueron seleccionados para el análisis multivariado, es de mencionar, que con este análisis se pretende analizar de manera simultánea el efecto conjuntos de estos factores que fueron seleccionados mediante el análisis de componentes principales realizado a cada categoría estudiada en esta investigación. Así mismo la importancia de este análisis multivariado radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que a través de los métodos estadísticos univariados y bivariados sería casi imposible de conseguir.

Factores para el análisis multivariado

- ✓ Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto (Capacidad de los desarrolladores)
- ✓ Amplios conocimientos de los desarrolladores sobre los aspectos de negocio del producto a desarrollar (Conocimiento del dominio)
- ✓ Gran capacidad de comunicación de los desarrolladores (Comunicación)
- ✓ Excelentes conocimientos técnicos de los diseñadores (Composición del equipo)
- ✓ Tamaño del equipo del proyecto (Composición del equipo)
- ✓ Cantidad de interfaces externas del producto tanto de hardware como software (Complejidad del producto)
- ✓ Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto (Madurez de gestión de proyectos)
- ✓ Nivel de continuidad y estabilidad del equipo del proyecto (Colaboración del equipo)

Para esta parte final del análisis se tuvieron en cuenta aquellos factores donde los niveles críticos no se encuentre marcados en azul en la tabla 8, es decir no presenten correlaciones estadísticamente significativas con otros factores, o en su defectos que dichas correlaciones no sean suficientemente fuertes (evitar multicolinealidad), en este sentido las factores que presentan estas características fueron divididos en dos grupos: Tamaño del equipo del proyecto (Composición del

equipo) y Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto (Madurez de gestión de proyectos), ó Cantidad de interfaces externas del producto tanto de hardware como software (Complejidad del

producto), y Excelentes conocimientos técnicos de los diseñadores (Composición del equipo), los posteriores análisis ayudo a determinar cual grupo de variables seleccionar.

MATRIZ DE CORRELACIONES Rho de SPEARMAN		Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto	Amplios conocimientos de los desarrolladores sobre los aspectos de negocio del producto a desarrollar	Gran capacidad de comunicación de los desarrolladores	Excelentes conocimientos técnicos de los diseñadores	Tamaño del equipo del proyecto	Cantidad de interfaces externas del producto (tanto de hardware como software)	Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto	Nivel de continuidad y estabilidad del equipo del proyecto
Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto		.390**	.435**	.508**		-.019	.076	.428**	.451**
		.000	.000	.000		.868	.493	.000	.000
Amplios conocimientos de los desarrolladores sobre los aspectos de negocio del producto a desarrollar			.268*	.367**		-.150	.056	.383**	.209
			.014	.001		.175	.616	.000	.058
Gran capacidad de comunicación de los desarrolladores				.284**		-.104	.011	.372**	.267*
				.009		.352	.921	.001	.015
Excelentes conocimientos técnicos de los diseñadores						-.062	.007	.739**	.172
						.580	.947	.000	.121
Tamaño del equipo del proyecto							.589**	-.120	.093
							.000	.282	.402
Cantidad de interfaces externas del producto (tanto de hardware como software)								-.042	.347**
								.705	.001
Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto									.199
									.071

Tabla 8 Matriz de correlaciones entre los factores seleccionados para la construcción de un modelo multivariado.

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.697
Prueba de esfericidad de Bartlett	Chi-cuadrado aprox	205,291
	gl	28
	Sig.	.000

Tabla 9 KMO y prueba de Bartlett

Por lo tanto, la prueba de Bartlett calculada fue igual a 0.697 alcanzando un nivel de regular, de acuerdo a la escala presentada, por tanto el modelo factorial empleado nos indica

que nos encontramos que las relaciones existentes entre estos factores son consideradas débiles.

En el tabla de comunidades que se presenta a continuación, se midió el porcentaje de la varianza de cada una de los factores que se explica por el resto de las factores que componen la categoría, en el caso específico de esta investigación el factor Cantidad de interfaces externas del producto (tanto de hardware como software) fue explicado en un 80% por los demás factores.

	Inicial	Extracción
Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto	1,000	.611
Amplios conocimientos de los desarrolladores sobre los aspectos de negocio del producto a desarrollar	1,000	.412
Excelentes conocimientos técnicos de los diseñadores	1,000	.665
Tamaño del equipo del proyecto	1,000	.669

Cantidad de interfaces externas del producto (tanto de hardware como software)	1,000	,803
Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto	1,000	,651
Gran capacidad de comunicación de los desarrolladores	1,000	,488
Nivel de continuidad y estabilidad del equipo del proyecto	1,000	,463

Tabla 10: Comunidades

El cuadro varianza total explicada muestra un modelo que es explicado por dos componentes. En la columna " Sumas de las saturaciones al cuadrado de la rotación " indica que el primer componente explica por sí solo el 37.2% de la variación total del grupo de factores seleccionados, mientras el segundo componente solo aporta el 22.2% de la variación

total de esta misma, en resumen ambos componentes juntos explican el 69.5% de la variación total.

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3.000	37.500	37.500	3.000	37.500	37.500	2.981	37.263	37.263
2	1.763	22.036	59.537	1.763	22.036	59.537	1.782	22.273	59.537
3	.911	11.385	70.922						
4	.714	8.926	79.848						
5	.571	7.136	86.983						
6	.482	6.023	93.006						
7	.310	3.881	96.887						
8	.249	3.113	100.000						

Tabla 11: Varianza total explicada

En la Matriz de componentes y Matriz de componentes de componentes rotados, se destaca que el factor Cantidad de interfaces externas del producto (tanto de hardware como software) presento en ambas tablas el valor más alto, sin embargo para este análisis nos centraremos en la Matriz de componentes de componentes rotados en ella podemos apreciar varias cosas, entre ellas que en el componente 1 los factores que la integraron con mayor fuerza fueron: Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto, Excelentes conocimientos técnicos de los diseñadores y Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto; por su parte, los factores más importante del componente dos fueron Tamaño del equipo del proyecto y Cantidad de interfaces externas del producto (tanto de hardware como software). En esta ocasión los factores que presentaron mayores valores no fueron incluidos para la construcción del modelo multivariado, esto se debe a que esos factores están más ampliamente relacionados con los demás factores y se puede incurrir en multicolinealidad, esto difiere a los análisis de episodios anteriores debido a que en esos momentos precisamente se necesitaba una factor que representara las categoría a la que pertenecía.

	Componente	
	1	2
Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto	,77	,115
Amplios conocimientos de los desarrolladores sobre los aspectos de negocio del producto a desarrollar	,63	-,089
Gran capacidad de comunicación de los desarrolladores	,68	-,120
Excelentes conocimientos técnicos de los diseñadores	,80	-,117
Tamaño del equipo del proyecto	-	,815
Cantidad de interfaces externas del producto (tanto de hardware como software)	,11	,889
Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto	,79	-,163
Nivel de continuidad y estabilidad del equipo del proyecto	,48	,482

Tabla 12: Matriz de componentes

	Componente	
	1	2
Amplio nivel de experiencia de los desarrolladores con la plataforma o entorno del proyecto	,753	,210

Amplios conocimientos de los desarrolladores sobre los aspectos de negocio del producto a desarrollar	,642	-,009
Gran capacidad de comunicación de los desarrolladores	,698	-,034
Excelentes conocimientos técnicos de los diseñadores	,815	-,016
Tamaño del equipo del proyecto	-	,800
	,172	
Cantidad de interfaces externas del producto (tanto de hardware como software)	,002	,896
Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto	,804	-,064
Nivel de continuidad y estabilidad del equipo del proyecto	,417	,538

Tabla 13: Matriz de componentes rotados

6. Despliegue.

La decisión de usar el modelo para ayudar a predecir el éxito o fracaso de un proyecto es decisión de la empresa, sin embargo, es conveniente incluir dentro de los procesos de la empresa la ejecución del modelo con el fin de que pueda dar una pequeña idea inicial de lo que se puede presentar en la ejecución de todo el proyecto.

III. CONCLUSIONES

El análisis de componentes principales permitió en primera instancia reafirmar la correlación existente entre los factores que componen cada categoría y segunda instancia reducir los factores que la componen, identificando el factor más relevante que debería ser seleccionado a la hora de realizar un análisis multivariado. Finalmente en el modelo multivariado permitió reducir el número de dimensiones necesarias que permitieran explicar el éxito o fracaso de un proyecto.

El modelo finalmente se puede utilizar como base para ayudar a predecir si un proyecto es exitoso o no, pero no puede ser considerado como única fuente de consulta, ya que no tiene capacidad de predicción en un 100%, esto debido a las variables y a la cantidad de datos utilizados, ya que hay muchísimos más factores que intervienen en el desarrollo de un proyecto los cuales no fueron considerados en el presente estudio.

A pesar de que los factores que se analizaron inciden en la inyección de defectos en el desarrollo de software según el estudio bibliográfico realizado, se puede determinar que los dos factores Excelente definición de roles y responsabilidades para las actividades de desarrollo del proyecto y Tamaño del equipo de proyecto si se les da un buen manejo y aplican el modelo respectivo, reduciría el número de inyección de defectos en el desarrollo de software, ya que por sí mismo ayudar a conseguir el éxito de los proyectos.

REFERENCIAS

- [1] J.L. Eveleens and C. Verhoef. (2010). The Rise and Fall of the Chaos Report Figures. *IEEE Software*, vol. 27, no. 1, pp. 30-36.
- [2] (2013) Blog Crhis F Carroll. [Online]. Available: <http://www.cafe-encounter.net/p1183/it-success-and-failure-the-chaos-report-factors>.
- [3] M.H. Nizam and S. Sahibuddin. (2011). Critical success factors for software projects: A comparative study. *Scientific Research and Essays*. Vol. 6(10), pp. 2174-2186. [Online]. Available: <http://www.academicjournals.org/sre/pdf/pdf2011/18May/Nasir%20and%20Sahibuddin.pdf>
- [4] K.R. Linberg. (1999). Software developer perceptions about software project failure: a case study, *Journal of Systems and Software*, Volume 49, Issues 2-3, 30, Pages 177-192. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0164121299000941>
- [5] R. J. Pereira, T. N. Cerpa and M. M. Rivas. (sa). Factores de éxito en proyectos de desarrollo de software: Análisis de la industria chilena del software. [Online]. Available: http://www.academia.edu/915974/Factores_de_exito_en_proyectos_de_desarrollo_de_software_analisis_de_la_industria_chilena_de_software.
- [6] M. Mas and C. Martínez. (2008). Análisis del factor estratégico para alcanzar el éxito de un proyecto de gestión del conocimiento. Aplicación al sector de la consultoría. *Dirección y organización, DY0 - 37*. [Online]. Available: <http://www.revistadyo.es/index.php/dyo/article/view/38>
- [7] Hernández Orallo, J. Ramírez Quintana, M.J and Ferri Ramírez, C. Introducción a la minería de datos. Madrid: Person Prentice Hall, 2004.
- [8] C. Cobos, J. Zúñiga, J. Guarín, E. León and M. Mendoza (2010). CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *Ingeniería e Investigación*, 30(3) 45-56. [Online]. Available: <http://www.redalyc.org/articulo.oa?id=64316140004>.
- [9] (2010) IBM Corporation. SPSS: Modeling: CRISP DM 1.0 - Step-by-step data mining guide. [Online]. Available: <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=YTW03084USEN>.
- [10] J. Jacobs, J.V. Moll, R. Kusters, J. Trienekens and A. Brombacher. (2007). Identification of factors that influence defect injection and detection in development of software intensive products, *Information and Software Technology*, Volume 49, Issue 7, Pages 774-789. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950584906001200>.

[11] M.A. Stutzke and C.S. Smidts. (2001). A stochastic model of fault introduction and removal during software development. IEEE Software, Volume 50, Issue 2, pp-184-193. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=963126&url=http%3A%2F%2Fieeexplore.ieee.org%2Fexpl%2Fabs_all.jsp%3Farnumber%3D963126

[12] X. Zhang and H. Pham. (2000). An analysis of factors affecting software reliability, Journal of Systems and Software, Volume 50, Issue 1, Pages 43-56. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0164121299000758>.