

# **Extracción de correlaciones entre el test de orientación vocacional de Chaside y la carrera de Ingeniería en Ciencias Informáticas**

## ***Extraction of correlations between Chaside's vocational orientation test and the degree in Computer Science Engineering***

### **Resumen:**

La sociedad cubana está comprometida con el éxito personal de cada uno de sus ciudadanos y el sistema de educación superior del país es un reflejo de ese compromiso. A pesar de ello los niveles de fracaso escolar no se corresponden con los deseados ya que sobrepasan los esperados. Una causa para esta problemática puede estar asociada a que los estudiantes no matriculan carreras correlacionadas favorablemente con sus habilidades, competencias y preferencias. En este trabajo se propone relacionar estas habilidades, competencias y preferencias presentes en el test de orientación vocacional CHASIDE con el rendimiento escolar en la carrera de Ingeniería en Ciencias Informáticas. El objetivo de este trabajo es obtener un modelo que permita hacer un análisis de la correlación que existe entre los elementos antes mencionados. Basándose en el proceso de desarrollo de software orientado al análisis de datos, o conocido como software de predominio de cómputo que usa como parte de su metodología la extracción de conocimiento de una base de datos, representado la información en forma de reglas de asociación que permiten obtener un modelo de análisis en forma de correlaciones. Se usó un dataset con un grupo de tuplas recogidas del test, para obtener el modelo que permite establecer las correlaciones entre las habilidades, competencias y preferencias y la carrera de Ingeniería en Ciencias Informáticas.

**Palabras clave: Competencias, correlación, habilidades, nivel de fracaso.**

**Abstract:**

*The Cuban society is committed to the personal success of each of its citizens and the country's higher education system is a reflection of that commitment. In spite of this, the levels of school failure do not correspond to the desired ones as they surpass those expected. A cause for this problem may be associated with students not enrolling in careers correlated favorably with their skills, competencies and preferences. In this work, proposes to relate these skills, competencies and preferences present in the orientation test vocational CHASIDE with the career of Engineering in Computer Science. The objective of this work is to obtain a model that allows an analysis of the correlation that exists between the elements aforementioned. Based on the process of software development oriented to the analysis of data, or known as computer predominance software that uses as part of its methodology the extraction of knowledge from a database, represented the information in the form of association rules that allow to obtain a model of analysis in the form of correlations. A dataset with tuples was used to obtain the model that allows to establish the correlations between the skills, competences and preferences and the career of Engineering in Computer Science.*

**Keywords: Competencies, correlation, skills, level of failure.**

## Introducción

La orientación vocacional comienza a considerarse un elemento importante en el desarrollo de capacidades creadas en los trabajadores a partir del año 1908 con la creación, en Boston, Estados Unidos, del Primer Buró de Orientación Vocacional a cargo de Frank Parsons. En aquel momento se introduce el término “Vocational Guidance” para agrupar un conjunto de elementos que permitían escoger qué profesión podría resultar más adecuada para cada persona en particular (Di Doménico, 2000).

A medida que la orientación vocacional ha evolucionado se han utilizado técnicas que permiten establecer determinadas correspondencias entre las aptitudes naturales del ser humano y competencias necesarias para el correcto desempeño de la profesión. Las técnicas que han acaparado la mayor atención en este tipo de tareas están asociadas a la aplicación de test que siguen las teorías factorialitas (De Fruyt, F. & Mervielde, 1997).

Cuba ha dedicado esfuerzos en el área de la orientación vocacional. Este es un proceso que se lleva a cabo en el sector de la educación, y que sirve de ayuda para que los estudiantes de enseñanza media y media superior puedan seleccionar una profesión. Sin embargo, a pesar de ello, los niveles de fracaso escolar en las universidades cubanas todavía alcanzan cuotas superiores a las esperadas (Maura, V. G, 2013). La insuficiente orientación vocacional es uno de los factores que provoca que los estudiantes seleccionen carreras para las que no tienen las aptitudes necesarias o que coinciden con sus intereses básicos. Esto provoca limitar la capacidad de los estudiantes para enfrentar las tareas, ya que no están motivados y les dificulta que alcancen resultados positivos.

Se han desarrollado varios instrumentos para determinar las habilidades, competencias y preferencias de los estudiantes y establecer la relación entre estas y las áreas del conocimiento. Especialmente interesante en este sentido es el test

CHASIDE (Maura, V. G,2013). Este test sirve de apoyo para identificar los intereses y aptitudes de los estudiantes enfocados a una orientación vocacional y profesional, el test consta de 98 preguntas con actividades que pueden ser del interés del estudiante, y así poder tener una guía de los intereses que el estudiante podría tener (ver anexo A). Sin embargo, estos resultados no han sido correlacionados, de manera experimental, con el éxito en el estudio de la carrera de Ingeniería en Ciencias Informáticas. Lo que constituye una limitación importante si se pretenden desarrollar sistemas inteligentes y/o de recomendación que contribuyan a la orientación de los estudiantes con vistas a decidir la carrera más favorable para ellos.

Lo antes expresado con anterioridad conduce a plantear que no existe un modelo a partir del test Chaside que resuelva: ¿Cómo correlacionar las habilidades, competencias y preferencias determinadas mediante el test de orientación vocacional CHASIDE con la carrera de Ingeniería en Ciencias Informáticas? Dicha solución se enmarcará en el proceso de desarrollo de software, pretendiendo: establecer un grupo de correlaciones entre las habilidades, competencias y preferencias detectadas en el test vocacional CHASIDE con la carrera de Ingeniería en Ciencias Informáticas.

## Metodología

Teniendo en cuenta que el modelo a representar responde a la necesidad de la Universidad de las Ciencias Informáticas de mejorar los resultados docentes alcanzados, especialmente en la disminución del fracaso escolar. La figura 1 pretende explicar cómo darle solución a unos de los problemas fundamentales por lo cual ocurre el fracaso escolar. El modelo de Reglas de Asociación que se pretende construir para identificar los estudiantes con aptitudes e intereses a fines a la carrera, permitirá combatir la mala orientación vocacional que pueda tener un estudiante hacia la carrera.

Luego de un estudio detallado de los conceptos básicos de los tipos de software, específicamente los de predominio de cómputo, se decide implementar una herramienta que permita realizar el proceso de extracción de conocimiento en bases de datos. Esta se realiza mediante el uso de los modelos de reglas de asociación, para obtener como resultado un modelo que describa las aptitudes e intereses presentes en el ingeniero en Ciencias Informáticas. El proceso de extracción de conocimiento se realizará mediante la metodología Catalyst (Pyle, D, 2003), la cual describe detalladamente los pasos a realizar y se ajustan las tareas a las características del problema. Específicamente centrándose en el escenario 1 ya que no se tiene referencias de que antes se allá intentado correlacionado los intereses y aptitudes con el desempeño académico en la carrera de Ingeniería en Ciencias Informáticas.

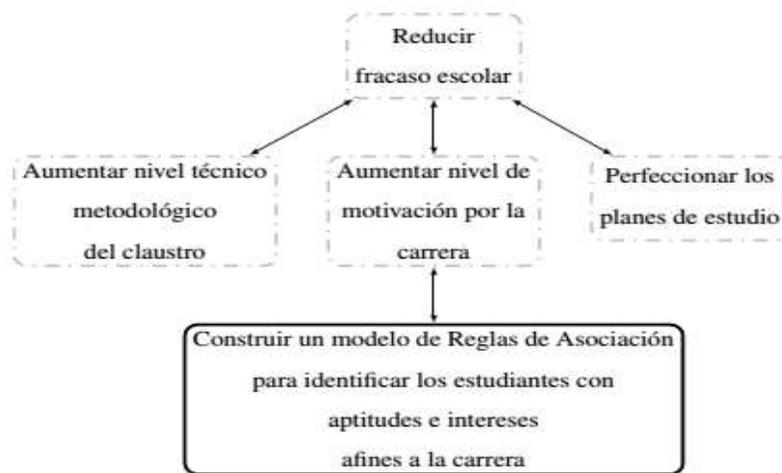


Figura 1. Relación del modelo de reglas con los objetivos de la institución.

### Obtención de los datos a explorar

Los datos necesarios para realizar la exploración se obtienen del test de orientación vocacional CHASIDE y la clasificación de los estudiantes se realiza de acuerdo a sus resultados en la carrera como sigue a continuación:

- Rendimiento normal índices entre 3.16 y 3.99.
- Buen rendimiento para valores entre 4 y 4.75.
- Alto rendimiento promedio por encima de 4.75.

Estas clasificaciones responden a la intención de un primer análisis exploratorio y no pretenden ser un criterio único para determinar el éxito de un estudiante en la carrera, esta fue definida a partir del criterio de un experto.

Se implementó una aplicación Web para recopilar los datos, vista principal se muestra en la figura 2, donde se responde las preguntas referentes al test CHASIDE.



- id\_f\_estudiantes: este campo identifica a cada encuestado.
- nombre: este campo almacena el Nombre del encuestado
- email:este campo almacena el correo del encuestado.
- centro\_de\_procedencia: este campo almacena la procedencia de cada encuestado.
- f\_provinciaid\_f\_provincia: este campo almacena el identificador de la provincia.
- id\_clasificacion: este campo almacena el identificador de la clasificación dada al encuestado y solo va a estar lleno para los encuestados de interés a ser analizados.



Figura 4. Diagrama físico de la tabla: f\_estudiantes

2. Tabla Clasificacion, figura 5, cuenta con los siguientes campos:

- id\_clasificacion: este campo identifica las cuatro clasificaciones posibles.
- Descripcion: este campo describe la clasificación dada.
- rangodeevaluacion: este campo almacena los principales rasgos por lo cual se dio la clasificación.

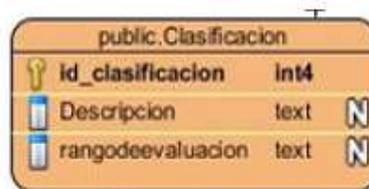


Figura 5. Diagrama físico de la tabla: Clasificacion

3. Tabla f\_fuente figura 6, cuenta con los siguientes campos:

-id\_f\_fuente: este campo identifica de cada pregunta por cada encuestado la respuesta dada.

-f\_estudiantesid\_f\_estudiantes: este campo hace referencia al encuestado.

-f\_preguntasid\_f\_preguntas: este campo hace referencia a la pregunta que marcó o no el encuestado.

-marcada: este campo almacena 1 si el encuestado marcó la pregunta o 0 si no.

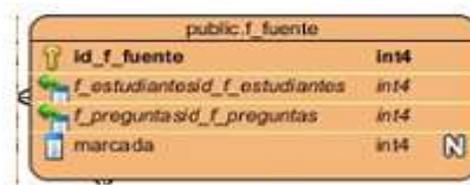


Figura 6. Diagrama físico de la tabla: `f_fuente`

4. Tabla `f_pregunta` figura 7, cuenta con los siguientes campos:

-id\_f\_preguntas: este campo identifica a la pregunta del test CHASIDE.

-descripcion: este campo describe la pregunta.

-f\_categoria\_de\_la\_carreraid\_f\_categoria\_de\_la\_carrera: este campo hace referencia a la categoría de la pregunta dentro del test CHASIDE.

-f\_tipo\_testid\_f\_tipo\_test: este campo hace referencia al tipo de la pregunta dentro del test CHASIDE la cual puede ser 1 para aptitud o 2 para interés.

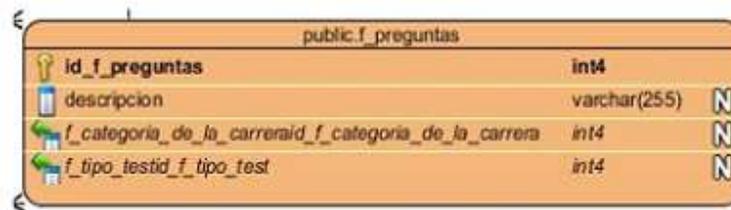


Figura 7. Diagrama físico de la tabla: `f_preguntas`

### Extracción de los datos desde las fuentes

Los datos a explorar, que constituyen los valores de las variables del problema, se obtiene mediante una función SQL. En la misma se genera un registro para cada estudiante que realizó el test y se le asocia el valor (true o false) a cada pregunta

del test. La representación final contendrá en cada fila los números de las preguntas marcadas en true. El total de preguntas es de 98 más un atributo en el que se refleja una categoría asociada al aprendizaje.

Después del análisis exploratorio de los datos se puede afirmar que no existen datos ausentes debido a que todas las preguntas fueron contestadas por cada encuestado. El 100 % de los 97 encuestados están clasificados de acuerdo a su promedio de notas.

La diversidad de la distribución de las clases es homogénea, hay un 35 % clasificados de “Avanzado”, un 33 % clasificados de “Satisfactorio” y un 32 % evaluado de “Bueno”. Las principales variables que son las clasificaciones de los encuestado y las preguntas están lo suficientemente completas para poder seleccionar la información deseada. Lo próximo que se debe hacer es verificar en qué estado está la información que se va a extraer.

También tras el análisis exploratorio se puede afirmar que:

- variables con un único valor en el 95 % de las instancias son: 50,69,8,67,74,21,29,7,63
- no aparecen variables con datos ausentes en las instancias.
- no aparecen variables con muchas categorías únicas.

El análisis exploratorio de las variables arroja que para la construcción del modelo no se van a tener en cuenta las variables con valor único ya que siempre van a estar incluidas en todas las reglas. Reduciendo así el análisis de las preguntas a las que tienen una ocurrencia por los encuestados menor que 1.

### **Chequear problemas en el conjunto de datos**

Al analizar los datos no se encontraron patrones inesperados durante la fase exploratoria, por lo que no es necesario una limpieza de los datos. Las preguntas 50, 69, 8, 67, 74, 21, 29, 7 y 63 siempre tienen el mismo valor, por ello se

consideran variables irrelevantes para el modelo ya que van a ser fijas en las reglas, no tienen peso para la construcción del modelo.

Tras dividir el conjunto de datos en diferentes partes para observar el comportamiento de las variables, se ha evidenciado que este se mantiene con respecto a la estructura, por lo cual se concluye que los datos son suficientes para la clasificación y están correctamente estructurados. Afirmando que se puede proceder a realizar la preparación y modelado de los datos.

Lo dicho anteriormente hace que se decida implementar un algoritmo que permita controlar los parámetros de uso memoria y almacenamiento, que son problemas comunes que surgen en el proceso de datos con gran cantidad de columnas. Además, que permita procesar datos que no sean binarios para lo cual los datos se estructuran de la manera siguiente, para que puedan ser procesados por el algoritmo que se implementará.

Estructura de los datos:

- √ Preguntas de tipo aptitud (identificador).
- √ Preguntas de tipo interés (identificador).
- √ Clasificador.

La estructura planteada para el procesamiento de los datos cuenta con 98 preguntas que se representan de manera numérica por el identificador de la pregunta y además se le agrega una columna para poder clasificar a cada encuestado se hace necesario buscar un mecanismo capaz de mejorar este problema. Teniendo en cuenta los resultados esperados se caracterizan las variables de entrada y de salida para una mejor comprensión.

Teniendo como entrada la estructura anteriormente definida, se caracterizan las variables que la conforman. Al ser las preguntas parte de estas variables, se decide agruparlas por perfiles, teniendo en cuenta la clasificación que plantea el test CHASIDE (tomando como rango para el perfilado las columnas de cada

categoría). Los perfiles se hallaron al recorrer las preguntas de cada encuestado, tomando la lista de preguntas por cada rango definido y comparándolas con la lista de perfiles creados, en caso de no ser iguales, se crea un perfil con dicha lista. Al perfilar los datos, se reduce el número de columnas a 15, 7 perfiles de tipo aptitud, 7 perfiles de tipo interés y la clasificación del encuestado, así se garantiza que hallan menos filas para procesar.

En la solución se van a omitir todas aquellas preguntas o variables que tengan un valor único o esté su soporte por encima del 95 %. La variable de salida es un conjunto de reglas de asociación, el cual va estar conformado por la combinación de los perfiles que tienen en común cada clase o clasificación que va a implicar a dicha clasificación. La variable de salida se va a visualizar en forma de reglas de asociación, la cual debe tener un soporte y una confianza (Ejemplo: {Conjunto\_de\_perfiles(combination\_de\_perfiles)} → clasificador, Sop:X, Conf : Y , o sea las preguntas 69,214,215,14,48→3(clasificador),Confianza:0.571414,Soporte:1.0). El resultado esperado es un modelo de reglas de asociación que permite correlacionar el comportamiento académico de los estudiantes en la carrera de Ingeniería en Ciencias Informáticas.

El modelo será implementado como parte del sistema donde están almacenados los datos del test CHASIDE, para que sea posible hallar la correlación entre el test con otras carreras, convirtiéndose en una herramienta adaptable para el estudio de cualquier carrera. Con esta información se podrá predecir el comportamiento de un egresado del nivel medio superior en una carrera donde se halla echo el estudio, en este caso para la en particular para la carrera de Ingenierías en Ciencias Informáticas. El resultado también servirá para evaluar las competencias e intereses de los profesionales en una carrera.

### **Construcción del modelo**

Como se ha establecido anteriormente lo primero que se necesita es seleccionar los datos con que se va a trabajar para luego crear la vista minable. Teniendo en

cuenta que se necesitan todos los perfiles por cada tipo de test dígame interés y aptitud con la clasificación dada a cada encuestado, se construyen 2 funciones para poder extraer los datos y transformarlos en la vista minable que se necesita. Luego de estructurar los datos se implementa el algoritmo Apriori en Python el cual tiene como entrada la vista minable, el soporte mínimo, la confianza mínima y los items, devolviendo como salida una lista de reglas de asociación la cual se guardan en la base de datos (Zaki, M. J,2014).

La estructura de la vista minable creada está compuesta por una fila que constituye un encuestado con los perfiles 7 para aptitud, 7 para interés y la última columna hace referencia a la clasificación. A continuación, se describe la estructura:

C	H	A	S	I	D	E	C	H	A	S	I	D	E	Clase
Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Perfil	Clasificación
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Intereses							Aptitudes							

Una vez ejecutado el modelo con un soporte mayor a 0.4 y una confianza mayor de 0.3 se obtienen 295 reglas de las cuales 52 reglas clasifican a la categoría de satisfactorio, 64 reglas clasifican a la categoría de bueno y 179 reglas clasifican a la categoría de avanzado. De las reglas encontradas al ejecutar el proceso se muestra a un conjunto con algunas de ellas en la tabla 1:

Tabla 1. Modelo de Reglas de asociación.

Lista de preguntas	Clasificador	Sopрте	Confianza
0,12,17,43	Satisfactorio	0.0909091	0.666667
9,59,0	Satisfactorio	0.0909091	0.666667
72,14,38,46,51,0	Satisfactorio	0.0909091	0.666667
72,0	Satisfactorio	0.0909091	0.666667
0,17,43	Satisfactorio	0.136364	0.6
3,28,40,0	Satisfactorio	0.136364	0.5
3,0	Satisfactorio	0.0909091	0.5
72,14,38,46,51	Satisfactorio	0.0909091	0.5
17,43,5,18,37,47,54,60,83	Satisfactorio	0.0909091	1
3,9,59	Satisfactorio	0.0909091	1
24,33,41,56,80,89,95,76,82	Bueno	0.0909091	1
24,33,41,56,80,89,95,72,86,76,82	Bueno	0.0909091	1
9,59,39,82	Bueno	0.0909091	1
39,82	Bueno	0.0909091	1
28,40	Bueno	0.0909091	1
16,31,34,42,49,61,68,77,88,93,28,40	Bueno	0.0909091	1
72,86,6,55,79	Bueno	0.0909091	1
6,55,79	Bueno	0.136364	0.75
14,38,46,51,9,25,59	Bueno	0.0909091	0.666667
15,32,44,52,70,87,92	Bueno	0.0909091	0.666667
6,55,79,94	Avanzado	0.136364	1
72,86,9,25,59,90	Avanzado	0.136364	1
6,55,79,94,72,86	Avanzado	0.136364	1
9,25,59,90	Avanzado	0.136364	1
14,38,46	Avanzado	0.0909091	1
14,38,51,9,25,59,90	Avanzado	0.136364	1
6,55,79,94,3,28,40	Avanzado	0.0909091	1
6,55,79,94,15,22,32,44,52,62,70,87,92	Avanzado	0.0909091	1
10,27,35,45,57,81,96	Avanzado	0.0909091	1
9,25,59,90,3,28,40	Avanzado	0.0909091	1

### Validación del modelo

La validación de un estudio es la cualidad que lo hace creíble y da testimonio del rigor con que se realizó. Esta implica la relevancia del estudio con respecto a sus objetivos, así como la coherencia lógica entre sus componentes (Yacuzzi, 2005).

Existen diversas técnicas para validar los modelos, como son la comparación de los parámetros con los obtenidos mediante modelos físicos teóricos o con simulaciones, utilizar nuevos conjuntos de datos conocidos para comparar con los obtenidos o el uso de técnicas de validación cruzada (Kozak, 2003), técnica que se utilizó en la presente investigación para validar los resultados. En particular la técnica de validación cruzada que se aplicará, va a permitir evaluar la solidez del

modelo de minería de datos, mediante el uso de la métrica de precisión, que se calcula de la siguiente forma:

$$\text{precisión} = \frac{\text{\#desaciertos}}{\text{\#objetos Clasificados}}$$

En la figura 8 se observa la estrategia de validación a seguir, hacerse en tres experimentos, donde se seleccionan dos clases para construir el modelo y se utiliza una tercera clase para clasificar la población que se estudia, o sea la tercera clase se emplea como criterio de clasificación.

Conjunto de datos de prueba (30 por ciento)	Conjunto de datos de entrenamiento (70 por ciento)
	<b>Partición 1</b> Cuando esta partición se utiliza como datos de prueba, el modelo se entrena en los datos en las Particiones 2 y 3.
	<b>Partición 2</b> Cuando esta partición se utiliza como datos de prueba, el modelo se entrena en los datos en las Particiones 1 y 3.
	<b>Partición 3</b> Cuando esta partición se utiliza como datos de prueba, el modelo se entrena en los datos en las Particiones 1 y 2.

Figura 8. Características de los datos para la validación cruzada con 3 clases.

Los experimentos se crean con tres clases de 32 encuestados cada uno, una construyendo el modelo con todas las reglas generadas. Los resultados obtenidos por cada caso fueron:

1. Experimento 1: Se tomaron la clase 1 y la clase 2 para construir el modelo, dejando la clase 3 para clasificar.

-El modelo tiene 192 reglas de asociación

-De los 32 casos que se usaron para clasificar, se clasificaron correctamente 29 casos

-Precisión =  $28/32$

2. Experimento 2: Se tomaron la clase 1 y la clase 3 para construir el modelo, dejando la clase 2 para clasificar. El modelo construido arrojó como resultado:

-El modelo tiene 203 reglas de asociación

-De los 32 casos que se usaron para clasificar, se clasificaron correctamente 30 casos

-Precisión =  $30/32$

3. Experimento 3: Se tomaron la clase 3 y la clase 2 para construir el modelo, dejando la clase 1 para clasificar. El modelo construido arrojó como resultado:

-Un modelo con 228 reglas de asociación

-De los 32 casos que se usaron para clasificar, se clasificaron correctamente 29 casos

-Precisión =  $29/32$

Las pruebas realizadas arrojaron como resultado final que, el promedio de precisión a partir del proceso de extracción de reglas de asociación, fue de un 89.33 % para el modelo completo. Concluyendo que modelo es estable y que tiene una presión a la hora de construir el modelo de un 89.33% de aceptación en la clasificación.

## Resultados y discusión

Al aplicar el proceso para extracción reglas de asociación al test de orientación vocacional CHASIDE con relación al desempeño académico, que se le aplicó a profesionales y estudiantes de la carrera se obtiene un modelo de reglas para cada combinación de los umbrales de soporte y confianza. A partir del total de reglas generadas se puede determinar qué combinaciones de intereses y aptitudes debe tener un estudiante con relación a su desempeño. El modelo planteado logra establecer un conjunto de 295 reglas las cuales describen características principales del ingeniero en Ciencias Informáticas. Logrando así correlacionar las aptitudes, habilidades y competencias del estudiante con el desarrollo académico en la carrera. El 56 % de las reglas encontradas tienen una confianza de 1 y el modelo construido a partir de la validación cruzada arrojó que para los 3 experimentos la precisión fue mayor que el 84 %, llegando a alcanzar un promedio de 89,6 %.

Un ejemplo de las reglas de asociación encontradas es el estudiante que tiene las aptitudes e intereses:

15,32,44,52,70,87,92,9,25,59,90→Avanzado con un soporte de 0.909091 y una confianza de 1. La combinación de dichas aptitudes e intereses son las características que debe tener un estudiante para ser avanzado en la carrera de Ingeniería en Ciencias Informáticas con una confianza de 1 y un soporte de 0,91 %.

### Leyenda

Las siguientes preguntas pertenecen a la categoría Medicina y Ciencias de la Salud dentro de los intereses de los estudiantes:

15- ¿Convences fácilmente a otras personas sobre la validez de tus argumentos?

32- ¿Participarías en una campaña de prevención de la enfermedad del Dengue?

44- ¿Te gustaría hacer un curso de primeros auxilios?

52- ¿Te resultaría interesante el estudio de las ciencias biológicas?

70- ¿Te gustaría investigar sobre una nueva vacuna?

87- Ante una emergencia epidémica ¿Participarías en una campaña brindando tu ayuda?

92- Ante un llamado solidario ¿Te ofrecerías para cuidar un enfermo?

Las siguientes preguntas pertenecen a la categoría Ingeniería y Computación dentro de las aptitudes de los estudiantes:

9- ¿Eres exigente y crítico con tu equipo de trabajo?

25- ¿Planificas detalladamente tus trabajos antes de empezar?

59- ¿Crees que tus ideas son importantes, y haces todo lo posible para ponerlas en práctica?

90- ¿Trabajar con objetos te resulta más gratificante que trabajar con personas?

## **Conclusiones**

Como resultado de la presente investigación se obtuvo un modelo para extraer reglas de asociación a partir del test de orientación vocacional CHASIDE para correlacionar las habilidades, intereses y competencias del estudiante de la carrera de Ingeniería en ciencias Informáticas con el desempeño académico. El modelo responde a la necesidad de la universidad de mejorar los resultados docentes alcanzados, especialmente en la disminución del fracaso escolar. En base a los resultados obtenidos se arribó a las siguientes conclusiones:

1. Se introduce la minería de datos como proceso para extraer conocimiento del test CHASIDE en forma de reglas de asociación, que también se utilizan para expresar el conocimiento previo. Esto se respalda mediante el uso de la metodología la cual describe detalladamente los pasos a realizar y se ajustan las tareas a las características del problema.

2. Se implementó una aplicación para la captura de los datos del test CHASIDE, lo que permitió su posterior procesamiento.
3. Se determinaron las técnicas de exploración de datos a utilizar mediante la metodología Catalyst utilizando la tarea descriptiva y la técnica de reglas de asociación. Logrando obtener un modelo de reglas de asociación el cual identificar las habilidades, intereses y competencias según el comportamiento académico del estudiante en Ciencias Informáticas.
4. Se implementa el algoritmo Apriori, una vez perfilados los datos para procesar la vista minable y extraer el modelo de reglas de asociación, el cual permitió establecer las aptitudes e intereses de los estudiantes en la carrera de Ingeniería en Ciencias Informáticas.
5. La realización pruebas de validación cruzada evidenció la precisión del modelo propuesto, mostrando que es preciso en un 89 %. Además, el modelo propuesto facilita la clasificación en tres clases (Rendimiento normal, Buen rendimiento y Alto rendimiento) de un encuestado teniendo en cuenta las preguntas marcadas.

## **Bibliografía**

- [1] Di Doménico, C. & Vilanova, A. Orientación vocacional: origen, evolución y estado actual. *Orientación y sociedad*, vol. 2, pp. 47–58, 2000. Consultado el: 2017-06-08.
- [2] De Fruyt, F. & Mervielde, I. The five-factor model of personality and Holland's RIASEC interest types. vol. 23, *Personality and Individual Differences*, no. 1, pp. 87–103, Julio 1997. Consultado el: 2016-11-10.
- [3] Maura, V. G. El servicio de orientación vocacional-profesional (SOVP) de la Universidad de La Habana: una estrategia educativa para la elección y desarrollo profesional responsable del estudiante. vol. 6, *Pedagogía Universitaria*, no. 4, 2013. Consultado el: 2016-11-09.
- [4] Pyle, D. *Business modeling and data mining*. Morgan Kaufmann, 2003.

