



UNIVERSIDAD CATÓLICA DE SALTA
FACULTAD DE INGENIERÍA
INGENIERÍA EN INFORMÁTICA

***Minería de datos aplicada a un sistema de
alarma para la detección del riesgo
académico en un colegio secundario.***

Proyecto de Grado.

Alumna: Rodríguez Estefania Janet.

2019

Título

Ingeniería en Informática

Profesor Guía

Nombre: Lic. Cardoso Carolina

Firma: _____

Tribunal Evaluador:

Nombre: Ing. Rivadera Gustavo

Firma: _____

Nombre: Lic. Perdiguero Jorge

Firma: _____

Nombre: Ing. Diaz Carlos

Firma: _____

Alumna

Nombre: Rodriguez Estefania.

Firma: _____

Fecha de exposición:

__/__/__

Dedicatoria y agradecimientos.

Quiero agradecer principalmente a Dios por ser mi guía y fortaleza en todos los momentos de mi vida.

A mi madre Fanny, quien con tanta paciencia y amor me acompañó y ayudó en los mejores y peores momentos de mi vida.

A mi padre Gustavo por todo el apoyo y ayuda brindada desde siempre.

A mi hermana Joana, quien siempre me acompaña y quien con tanta paciencia y amor también me ha acompañado en cada paso realizado.

¡GRACIAS FAMILIA POR TANTO AMOR!

Índice

| | |
|--|----|
| Resumen | 1 |
| Introducción..... | 2 |
| Capítulo 1: Estado de la cuestión..... | 4 |
| 1.1. Dato, información y conocimiento | 4 |
| 1.2. Minería de datos o Data Mining | 5 |
| 1.2.1 La minería de datos y el proceso KDD | 5 |
| 1.2.2. Modelos de la minería de datos..... | 9 |
| 1.2.3. Tareas de la minería de datos | 10 |
| 1.2.4. Técnicas de minería de datos..... | 10 |
| 1.3. Metodología CRISP-DM | 12 |
| 1.4. Sistema | 16 |
| 1.5. Rendimiento académico..... | 16 |
| 1.6. Riesgo académico | 18 |
| 1.7. Repitencia escolar | 19 |
| 1.8. Investigaciones previas | 19 |
| Capítulo 2: Descripción del problema. | 21 |
| 2.1. Objetivo General | 22 |
| 2.1.2. Objetivos específicos..... | 22 |
| 2.2. Alcance | 22 |
| 2.3. Alternativas tecnológicas | 23 |
| Capítulo 3: Solución propuesta..... | 25 |
| 3.1. Herramientas utilizadas..... | 25 |
| 3.2. Desarrollo de la solución..... | 25 |
| <input type="checkbox"/> Fase 1: Entendimiento del negocio | 25 |
| <input type="checkbox"/> Fase 2: Entendimiento de datos:..... | 30 |
| <input type="checkbox"/> Fase 3: Preparación de los datos:..... | 36 |
| <input type="checkbox"/> Fase 4: Modelado:..... | 41 |
| <input type="checkbox"/> Fase 5: Evaluación: | 42 |
| <input type="checkbox"/> Fase 6: Despliegue: | 43 |
| 3.3. Funcionamiento interno de la solución propuesta | 43 |
| 3.4. Análisis FODA | 44 |
| 3.4.1. Fortalezas | 44 |

| | |
|--|----|
| 3.4.2. Debilidades | 44 |
| 3.4.3. Oportunidades | 44 |
| 3.4.4. Amenazas..... | 44 |
| 3.5. Análisis de factibilidad | 45 |
| 3.5.1. Factibilidad económica..... | 45 |
| 3.5.2. Factibilidad técnica | 47 |
| 3.5.3. Factibilidad legal..... | 47 |
| 3.6. Análisis de riesgos | 47 |
| Capítulo 4: Pruebas y resultados..... | 49 |
| 4.1. Prueba N° 1..... | 50 |
| 4.2. Prueba N° 2: Operador Chi-squared. | 52 |
| 4.3. Prueba N° 3: Operador Remove Useless Attributes. | 55 |
| 4.4. Incorporación del modelo obtenido al sistema de alarma..... | 59 |
| Conclusión..... | 61 |
| Bibliografía..... | 63 |
| Anexo I. Sistema de alarma..... | 66 |
| Anexo II. Alarma del sistema- Código. | 73 |
| Anexo III. Instalación del sistema de alarma. | 77 |

Índice de imágenes.

| | |
|--|----|
| Imagen 1-Dato e Información..... | 4 |
| Imagen 2- Proceso KDD | 6 |
| Imagen 3- Matriz de confusión..... | 7 |
| Imagen 4- Validación cruzada..... | 9 |
| Imagen 5-Fases del proceso de minería de datos. (Shearer, 2000)..... | 13 |
| Imagen 6- Fase 2- Minería de datos..... | 14 |
| Imagen 7-Fase 3- Minería de datos..... | 15 |
| Imagen 8-Modelo de encuesta (Parte 1)..... | 26 |
| Imagen 9-Modelo de encuesta (Parte 2)..... | 27 |
| Imagen 10-Modelo de encuesta (Parte 3)..... | 27 |
| Imagen 11-Modelo de Encuesta (parte 4) | 28 |
| Imagen 12-Registro de calificaciones | 29 |
| Imagen 13-Registro de asistencias..... | 30 |
| Imagen 14-Excel generado (parte 1)..... | 31 |
| Imagen 15-Excel generado (parte 2)..... | 31 |
| Imagen 16-Excel generado (parte 3)..... | 32 |
| Imagen 17-Importando archivo Excel..... | 33 |
| Imagen 18-- Importando archivo Excel (selección de tipo de dato)..... | 33 |
| Imagen 19-Cantidad de repitentes y no repitentes..... | 34 |
| Imagen 20-Histograma para atributo EDAD..... | 35 |
| Imagen 21-Edad de alumnos repitentes | 35 |
| Imagen 22-Histograma para atributo PASE..... | 36 |
| Imagen 23-Campos con más de un valor | 37 |
| Imagen 24- Valores dispersos para campo CANT. DE HERMANOS..... | 38 |
| Imagen 25-Valores dispersos para campo TIEMPO DEDICADO A LOS ESTUDIOS..... | 39 |
| Imagen 26-Histograma para atributo clase (RIESGO ACADÉMICO)..... | 40 |
| Imagen 27-Proceso de validación cruzada | 42 |
| Imagen 28-Funcionamiento interno de la solución..... | 43 |
| Imagen 29-Proceso de modelado Cross Validation..... | 49 |
| Imagen 30-Subproceso cross validation | 49 |
| Imagen 31-Árbol de decisión obtenido | 50 |
| Imagen 32-Matriz de confusión del árbol de decisión..... | 50 |
| Imagen 33-Matriz de confusión del algoritmo Naive Bayes..... | 51 |
| Imagen 34-Matriz de confusión del algoritmo WJ48 | 51 |
| Imagen 35- Descripción de algoritmo WJ48..... | 51 |
| Imagen 36-Matriz de confusión del algoritmo W-OneR | 52 |
| Imagen 37-Descripción de algoritmo W-OneR..... | 52 |
| Imagen 38--Proceso con CHI-SQUARED..... | 53 |
| Imagen 39- W-OneR con el operador CHI-SQUARED..... | 53 |
| Imagen 40-Descripción de algoritmo W-OneR..... | 53 |
| Imagen 41-Matriz de confusión de Decision Tree con el operador CHI-SQUARED.... | 54 |
| Imagen 42-Árbol de decisión | 54 |
| Imagen 43-Matriz de confusión de árbol de decisión con el operador CHI-SQUARED54 | |
| Imagen 44-Descripción de algoritmo WJ48..... | 55 |

| | |
|--|----|
| Imagen 45- Matriz de confusión de Naive Bayes..... | 55 |
| Imagen 46-Proceso con operador REMOVE USSLES | 55 |
| Imagen 47-Árbol de decisión obtenido con REMOVE USSLES | 56 |
| Imagen 48-Matriz de confusión de árbol de decisión con operador REMOVE USSLES | 56 |
| Imagen 49-Matriz de confusión de WJ48 con operador REMOVE USSLES | 56 |
| Imagen 50-Descripción de algoritmo WJ48..... | 57 |
| Imagen 51-Matriz de confusión de W-OneR con operador REMOVE USSLES | 57 |
| Imagen 52-Descripción de algoritmo W-OneR..... | 57 |
| Imagen 53- Matriz de confusión de Naive Bayes con operador REMOVE USSLES ... | 57 |
| Imagen 54-Matriz de confusión con precisión más alta- Naive Bayes..... | 58 |
| Imagen 55-Proceso que utiliza el modelo generado y un nuevo conjunto de datos para predecir nuevas instancias..... | 59 |
| Imagen 56-Resultados..... | 59 |
| Imagen 57- Página Principal | 66 |
| Imagen 58- Inicio de Sesión..... | 67 |
| Imagen 59- Menú principal | 68 |
| Imagen 60- Carga de datos (parte 1)..... | 68 |
| Imagen 61- Carga de datos (parte 2)..... | 69 |
| Imagen 62- Cargar inasistencias..... | 69 |
| Imagen 63- Cargar notas | 70 |
| Imagen 64-Visualización del riesgo académico | 71 |
| Imagen 65-Visualización de datos personales y académicos de un alumno en particular | 72 |
| Imagen 66- Inicio de RapidMiner..... | 77 |
| Imagen 67-Instalando SQL Server. Paso 1 | 78 |
| Imagen 68-Instalando SQL Server. Paso 2 | 78 |
| Imagen 69-Instalando SQL Server. Paso 3 | 79 |
| Imagen 70-Instalando SQL Server. Paso 4 | 79 |
| Imagen 71-Instalando SQL Server. Paso 5 | 80 |
| Imagen 72-Instalando SQL Server. Paso 6 | 80 |
| Imagen 73-Instalando SQL Server. Paso 7 | 81 |
| Imagen 74-Instalando SQL Server. Paso 8 | 81 |
| Imagen 75-Instalando SQL Server. Paso 9 | 82 |
| Imagen 76-Instalando SQL Server. Paso 10 | 82 |
| Imagen 77-Instalando SQL Server. Paso 11 | 83 |
| Imagen 78-Instalando SQL Server. Paso 12 | 83 |
| Imagen 79-Copiando nombre del servidor..... | 84 |
| Imagen 80-Abriendo archivo CONFIG | 84 |
| Imagen 81- Modificando Data Source..... | 84 |

Índice de tablas

| | |
|---|----|
| Tabla 1- Modificación de atributos..... | 39 |
| Tabla 2- Atributos que conforman la vista minable | 41 |
| Tabla 3- Precisión y recall por algoritmo..... | 42 |
| Tabla 4- Herramientas de desarrollo y costos asociados | 45 |
| Tabla 5- Licencias Software..... | 45 |
| Tabla 6- Hardware necesario y costos asociados | 46 |
| Tabla 7- Recursos humanos necesarios y costos asociados | 46 |
| Tabla 8- Inversión total | 46 |
| Tabla 9- Análisis de riesgo..... | 48 |
| Tabla 10-Resumen de precisiones obtenidas en cada prueba..... | 58 |

Resumen

En este trabajo se desarrolló un sistema de alarma para la detección del riesgo académico basado en un modelo predictivo de minería de datos.

Para ello, fue necesaria la recolección de datos, la tabulación de los mismos, la aplicación de la minería de datos para obtener un modelo predictivo que logre identificar situaciones de riesgo escolar y los posibles factores influyentes en el rendimiento académico. Luego se diseñó el prototipo del sistema de alarma para incorporar dicho modelo.

Se utilizó la metodología de minería de datos denominada CRISP-DM.

La implementación de este sistema permite obtener información del estado escolar de los alumnos a los actores educativos como ser: preceptores y cuerpo directivo, y mediante ellos, también a los padres y/o tutores de los mismos.

El sistema de alarma basado en la minería de datos anticipa la situación académica de los alumnos en relación al bajo rendimiento escolar evitando de esta manera la interrupción de su trayectoria académica.

Introducción

El bajo rendimiento educativo y la repitencia escolar son dos problemas que tienen la mayoría de las instituciones educativas. Resolverlos se ha vuelto una tarea no muy sencilla debido a que su ocurrencia depende de múltiples factores. (Magaña Hernández, 2002)

Ante estas problemáticas y la preocupación que generan en el ámbito educativo, se realizaron múltiples investigaciones sobre la temática, presentando una variedad de posibles soluciones, entre ellas, una alternativa informática: la aplicación de la minería de datos. Ésta permite anticiparse a determinados hechos basándose en un conjunto de datos previamente analizados.

Aplicarla al ámbito escolar, es, en estos últimos tiempos de gran relevancia, debido a que es un proceso que permite identificar de manera temprana situaciones de riesgo académico en base a determinada información recolectada.

El colegio secundario N° 5091 Prof. Julio Cortes de la ciudad de San Ramón de la Nueva Orán, provincia de Salta, no queda exento a las problemáticas expresadas debido al incremento de alumnos repitentes y con bajo rendimiento.

En consecuencia, el presente trabajo desarrolla el prototipo de un sistema de alarma basado en la minería de datos con el fin de detectar el riesgo académico. Para ello es necesaria la aplicación de técnicas de recolección de datos, la tabulación de los datos obtenidos, la aplicación de la minería de datos para la obtención de un modelo predictivo e identificación de factores influyentes en el rendimiento académico y el desarrollo de interfaces que permitan visualizar la predicción del riesgo académico obtenido y a su vez que permita la interacción del usuario final con el sistema de alarma.

Con este sistema, la institución y sus actores educativos contarán con una herramienta informática que les permitirá tener información de la situación académica de los estudiantes en el momento que se requiera.

El presente trabajo aplica técnicas de minería de datos a un conjunto de datos provenientes del tercer año facilitados por la institución para obtener un modelo predictivo capaz de identificar anticipadamente situaciones de riesgo escolar.

La metodología seleccionada es la denominada CRISP-DM la cual describe el ciclo de vida del proceso de minería de datos en seis fases. Siguiendo ésta, se toman los datos obtenidos desde las técnicas de recolección de datos aplicadas para obtener un modelo que describa los posibles factores que pueden influir en el rendimiento académico.

Éste modelo incorporado al sistema de alarma, tiene múltiples alcances: los preceptores y el cuerpo directivo podrán acceder a la información personal y académica de cada alumno como así también podrán brindar información a los padres o tutores de los mismos de manera rápida y en el momento en que se lo requiera.

La información personal y académica de los alumnos será recolectada y cargada al sistema cada determinado tiempo.

El trabajo se estructura en 5 capítulos. En el capítulo 1 se detallan conceptos relevantes y trabajos relacionados a la temática. El capítulo 2 describe la problemática, los objetivos y el alcance del trabajo; luego el capítulo 3 describe la solución que se propone, mientras que en el capítulo 4 se detallan los resultados obtenidos y el capítulo 5 las conclusiones. Además, se concluye con tres anexos. El primer anexo detalla el funcionamiento del sistema de alarma. El segundo anexo contiene partes del código más importantes del sistema, mientras que el tercer anexo detalla los pasos a seguir para instalar el sistema.

Capítulo 1: Estado de la cuestión

1.1. Dato, información y conocimiento

Dos conceptos de suma importancia son nombrados a lo largo del presente trabajo de investigación, por ello, es necesario definirlos y explicarlos para poder comprenderlos.

El primero de ellos es el concepto de **dato** y el segundo es el concepto de **información**.

El termino **dato** es definido como “una representación de un objeto mediante el cual se pueden modelizar aspectos de un problema que se desea resolver con un programa sobre una computadora”. (Peter, 2006)

En otras palabras, los datos son las señales individuales en bruto y sin ningún significado que manipulan las computadoras para producir información.

Mientras que el termino **información** es definido como “un conjunto de elementos que permiten expresar conocimiento determinado”. Es decir, que la información es la interpretación y/o valoración que se le atribuye a un dato. Los datos se convierten en información cuando se relacionan con otros datos.

Por lo tanto, el dato tiene existencia real mientras que la información es de naturaleza lógica. (Peter, 2006)

Por último, el termino **conocimiento** es definido como la “capacidad de convertir datos e información en acciones efectivas”. (Daedalus, 2018)

En la Imagen 1 podemos visualizar de manera jerárquica la ubicación de cada uno de ellos:



Imagen 1-Dato e Información

Los dos primeros términos (dato e información) son considerados como sinónimos, pero en el ámbito informático no lo son y ambos están relacionados al último término (conocimiento) por su capacidad de conversión.

1.2. Minería de datos o Data Mining

Otro termino que requiere aclaración para la comprensión del presente trabajo es el termino *Minería de Datos*. Se define a la misma como “el proceso de extracción de conocimiento no trivial, comprensible, previamente desconocidos y potencialmente útil desde grandes cantidades de datos almacenados en distintos formatos”. (Hernandez Orallo, 2004)

El autor manifiesta que la principal tarea de la minería de datos es encontrar modelos comprensibles por medio de los datos. Cada patrón descubierto por medio de éste proceso debería ayudar en la toma de decisiones para que éstas sean más seguras y reporten un beneficio a la organización.

En este proceso se trabaja con grandes volúmenes de datos, procedentes de sistemas de información. Éstos, muchas veces, vienen acompañados de algunos problemas como ser: ruido, datos ausentes, intratabilidad, volatilidad de los datos, etc, por lo que es necesario usar técnicas adecuadas para analizarlos y extraer conocimiento novedoso y útil. (Hernandez Orallo, 2004)

1.2.1 La minería de datos y el proceso KDD

La minería de datos posee algunos sinónimos como el termino KDD (Knowledge Discovery in Databases) denominado como un proceso que consta de una serie de fases, donde la minería de datos es una de ellas. (Hernandez Orallo, 2004)

Asimismo, el termino KDD se define como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles, y en última instancia, comprensibles a partir de los datos”. (Fayyad, 1996)

En esta definición se resumen cuáles deben ser las propiedades deseables del conocimiento extraído:

- Válido: hace referencia a que los patrones deben seguir siendo precisos para dato nuevo y no solo para aquellos que hayan sido usados en su obtención.
- Novedoso: que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario.
- Potencialmente útil: la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.

- **Comprensible:** la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones.

Entonces se deduce que KDD es un proceso complejo que incluye no solo la obtención de los modelos o patrones (el objetivo de la minería de datos), sino también la evaluación y posible interpretación de los mismos, tal y como se refleja en la Imagen 2:



Imagen 2- Proceso KDD

Así, los sistemas KDD permiten la selección, limpieza, transformación y proyección de los datos, analizar los datos para extraer patrones y modelos adecuados, evaluar e interpretar los patrones para convertirlos en conocimiento, consolidar el conocimiento resolviendo posibles conflictos con conocimiento previamente extraído; y hacer el conocimiento disponible para su uso.

Esta definición del proceso clarifica la relación entre KDD y minería de datos: el KDD el proceso global de descubrir conocimiento útil desde las bases de datos mientras que la minería de datos se refiere a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos. (Hernandez Orallo, 2004)

El proceso KDD se organiza en cinco fases:

- 1- **Fase de integración y recopilación de datos:** en esta fase se determinan las fuentes de información que pueden ser útiles. Luego, se transforman y unifica la información recogida, detectando y resolviendo inconsistencias. Ya que los datos provienen de distintas fuentes, pueden contener datos faltantes o erróneos.
- 2- **Fase de selección, limpieza y transformación:** el siguiente paso es seleccionar y preparar el subconjunto de datos que serán útiles para crear una vista minable¹. Se realiza esto porque algunos datos coleccionados en la etapa

¹ Vista minable: es una tabla, la cual recoge toda la información necesaria (y solo esa) para realizar la minería de datos.

anterior pueden ser irrelevantes o anómalos, lo cual si no se realiza una selección y limpieza podemos conducir a resultados poco precisos.

Otra tarea de preparación de los datos es la creación de nuevos atributos a partir de los atributos que se tienen, con objeto de que estos hagan más fácil el proceso de minería.

3- **Fase de minería de datos:** el objetivo de esta fase es producir nuevo conocimiento que pueda utilizar el usuario. Aquí se construye un modelo basado en los datos recopilados. El modelo es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones pasadas.

4- **Fase de evaluación e interpretación:** será necesario explorar modelos alternativos hasta encontrar aquel que resulte más útil para resolver nuestro problema. En la búsqueda de “un buen modelo” puede que tengamos que retroceder hasta fases anteriores y hacer cambios en los datos que estamos usando o incluso modificar la definición del problema.

Aquí es necesario separar los datos en un conjunto de prueba y otro de entrenamiento para garantizar una precisión no muy sobreestimada.

La precisión se refiere a que tan cerca está el resultado de una medición del valor verdadero²; es una buena estimación de cómo se comportará el modelo para datos futuros.

Mediante lo que llamamos *matriz de confusión*, es posible observar la precisión calculada por cada algoritmo. La matriz de confusión es una tabla o matriz donde se observa los aciertos (Verdaderos Positivos y Verdaderos Negativos) y los no aciertos (Falsos Negativos y Falsos Positivos) que produce el algoritmo, como se muestra en la Imagen 3.

| | | Predicción | |
|-------------|-----------|---------------------------|---------------------------|
| | | Positivos | Negativos |
| Observación | Positivos | Verdaderos Positivos (VP) | Falsos Negativos (FN) |
| | Negativos | Falsos Positivos (FP) | Verdaderos Negativos (VN) |

Imagen 3- Matriz de confusión

- Los verdaderos positivos (VP) son la proporción de casos positivos que fueron correctamente identificados por el algoritmo.
- Los verdaderos negativos (VN) se trata de los casos negativos que el algoritmo ha clasificado correctamente.
- Los falsos negativos (FP) es la proporción de casos negativos que fueron erróneamente clasificados como positivos por el algoritmo.

² Precisión: <https://empresas.blogthinkbig.com/ml-a-tu-alcance-matriz-confusion/>

- Los falsos negativos (FN) es la proporción de casos positivos incorrectamente clasificados por el algoritmo.

La precisión se calcula de la siguiente manera:

$$\text{Accuracy (precisión): } (VP + VN) / (VP + VN + FP + FN).$$

El recall es la proporción de verdaderos positivos (VP) entre todos los casos positivos en los casos. Si su resultado es uno entonces se han encontrado verdaderos positivos (VP) en la base de datos, por lo que no existiría ruido ni silencio informativo. Por el contrario, si su valor es cero los datos no poseen relevancia alguna. (Haro Rivera, Zuñiga Lema, Freire Meneses, Vera Rojas, & Escudero Villa, 2018)

El recall se calcula de la siguiente forma³:

$$\text{Recall: } VP / (VP + FN)$$

Esto no garantiza que el modelo sea correcto, sino que simplemente indica que, si usamos la misma técnica con una base de datos con datos similares a los de prueba, la precisión media será bastante parecida a la obtenida con éstos. Existe otro método de evaluación denominado **validación cruzada** o **cross validation**, el cual es usado, generalmente, cuando son pocos los datos disponibles para construir el modelo o cuando no se dispone de un conjunto de prueba. Este método no es factible cuando se tienen conjuntos de datos grandes. (Hurtado, 2007)

Este método, divide los datos en **n** subconjuntos. Dado un subconjunto **k**, se usa **k** como prueba y los n-1 conjuntos restantes como entrenamiento. Esto se repite para cada subconjunto ⁴, como lo muestra la Imagen 4.

³ Calculo de Precisión y Recall:

<http://ceaa.esPOCH.edu.ec:8080/revista.perfiles/Articulos/Perfiles20Art13.pdf>

⁴ Método validación cruzada: <http://genome.tugraz.at/proclassify/help/pages/XV.html>

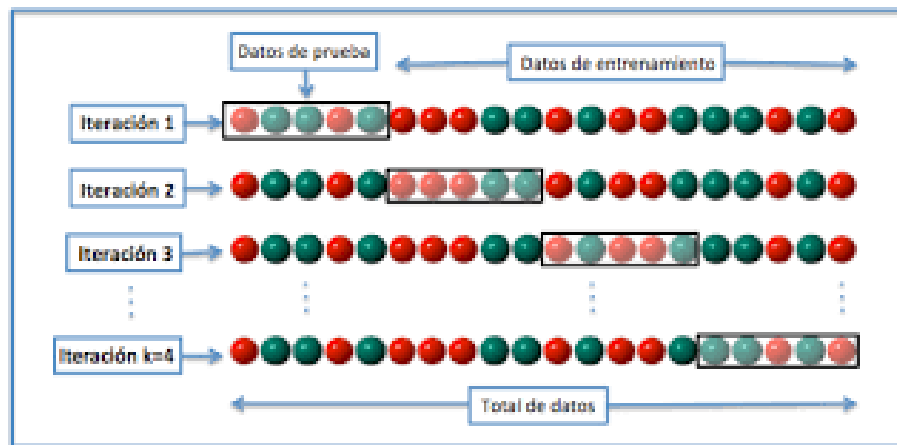


Imagen 4- Validación cruzada⁵.

- 5- **Fase de difusión, uso y monitorización:** en esta etapa se hace uso del conocimiento obtenido y es necesaria la difusión a los posibles usuarios. Además, es necesaria una cierta monitorización ya que los datos pueden ir variando. (Hernandez Orallo, 2004)

1.2.2. Modelos de la minería de datos.

La minería de datos tiene como objetivo analizar los datos para obtener conocimiento. Éste conocimiento puede ser en forma de relaciones, patrones o reglas inferidos de los datos y (previamente) desconocidos. Estas relaciones constituyen a su vez el modelo de los datos analizados. (Hernandez Orallo, 2004)

En la práctica estos modelos pueden ser de dos tipos:

- Modelo predictivo: pretenden estimar valores futuros o desconocidos de variables de interés que se denominan *variables dependientes*, usando otras variables o campos de las bases de datos que son denominadas *variables independientes*.
- Modelos descriptivos: exploran las propiedades de los datos examinados, no para predecir nuevos datos, sino para describirlos.

⁵ Gráfico de validación cruzada: https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada#cite_note-kfvcv-4

1.2.3. Tareas de la minería de datos

Las tareas de minería de datos pueden ser predictivas o descriptivas. Entre las tareas predictivas encontramos la clasificación y la regresión; mientras que en las de descripción tenemos: asociación y agrupamiento. (Hernandez Orallo, 2004)

La *clasificación* es la tarea más utilizada. En ella, cada registro (también llamado instancia) de la base de datos pertenece a una clase, la cual se indica mediante el valor de un atributo que llamamos la clase de la instancia. El objetivo es predecir la clase de las instancias de las que se desconoce la misma.

La *regresión* predice el valor de una variable continua basada en los valores de otras variables. Esta es la principal diferencia respecto a la clasificación, el valor a predecir es numérico. El objetivo en este caso es minimizar el error entre el valor predicho y el valor real.

El *agrupamiento (clustering)* consiste en obtener grupos a partir de los datos. Hablamos de grupos y no de clases, porque, a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, los analiza para generar esta etiqueta. Los datos son agrupados basándose en la similitud que tienen. Es decir, se forman grupos tales que los objetos del mismo grupo son similares entre sí, y al mismo tiempo, son muy diferentes a los objetos de otros grupos.

La *asociación* tiene como objetivo identificar relaciones no explícitas entre atributos *categoricos*. Es decir, dado un conjunto de registros, cada uno con elementos de alguna colección dada, produce reglas de dependencia que predigan la ocurrencia de un elemento basado en la ocurrencia de otros elementos. Este tipo de tarea se utiliza frecuentemente en el análisis de la cesta de la compra, para identificar productos que son frecuentemente comprados juntos, información esta que puede usarse para ajustar los inventarios o en campañas publicitarias. (Hernandez Orallo, 2004)

1.2.4. Técnicas de minería de datos

Las técnicas de minería de datos son algoritmos que se aplican sobre un conjunto de datos para obtener unos resultados. (Maimon & Rokach, 2010)

A continuación, se describen las técnicas de minería de datos más habituales:

- **Árboles de decisión:** el modelo obtenido tiene forma de árbol, es decir, son una serie de condiciones organizadas en forma jerárquica. Son muy útiles para encontrar estructuras en problemas que mezclen datos categoricos y numéricos. Los arboles de decisión siguen una aproximación “divide y vencerás” para partir el espacio del problema en subconjuntos. En el nodo raíz del árbol tenemos el problema a resolver. Los nodos internos corresponden a particiones sobre

atributos particulares y los arcos que emanan de un nodo corresponden a los posibles valores del atributo considerado en ese nodo. Cada arco conduce a otro nodo de decisión o a un nodo hoja. Los nodos hoja representan la predicción (o clase) del problema para todas aquellas instancias que alcanzan esa hoja. Para clasificar una instancia desconocida, se recorre el árbol de arriba hacia debajo de acuerdo a los valores de los atributos probados en cada nodo y, cuando se llega a una hoja, la instancia se clasifica con la clase indicada por esa hoja.

Los árboles de decisión pueden considerarse en forma de aprendizaje de reglas, ya que cada rama del árbol puede interpretarse como una regla, donde los nodos internos en el camino desde la raíz a las hojas definen los términos de la conjunción que constituye el antecedente de la regla, y la clase asignada en la hoja es el consecuente. (Hernandez Orallo, 2004)

Dentro de esta técnica, tenemos distintos algoritmos que permiten obtener árboles de decisión, entre ellos:

- W-J48: Se denomina algoritmo WJ48 a la implementación de Weka para el algoritmo C4.5.

El algoritmo C4.5 genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente. El árbol se construye mediante la estrategia de profundidad-primero. El algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información. Para cada atributo discreto, se considera una prueba con n resultados, siendo n el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos. En cada nodo, el sistema debe decidir cuál prueba escoge para dividir los datos. (Quinlan, 1993)

- Decisión Tree: es un árbol con una colección de nodos destinados a crear una decisión sobre la afiliación de valores a una clase o una estimación de un valor objetivo numérico. Cada nodo representa una regla de división para un atributo específico. La construcción de nuevos nodos se repite hasta que se cumplan los criterios de detención. Se determina una predicción para el atributo clase dependiendo de la mayoría de los ejemplos que alcanzaron esta hoja durante la generación, mientras que se obtiene una estimación de un valor numérico promediando los valores en una hoja.⁶

- Inducción de Reglas: es un conjunto de métodos para derivar un conjunto de reglas comprensibles de la forma:

SI cond ENTONCES pred.

El antecedente de la regla (la parte del SI) contiene una conjunción de n condiciones sobre los valores de los atributos independientes, mientras que el consecuente de la regla (la parte ENTONCES) contiene una predicción sobre el valor de un atributo objetivo. La semántica de este tipo de reglas de predicción es:

⁶ Obtenido de Documentación de RapidMiner. Descripción del operador Decision Tree.

si se satisfacen las condiciones especificadas, es decir **cond1, cod2,..., condn** en el antecedente de la regla para una instancia, entonces se predice que el atributo objetivo de este registro tendrá el valor especificado en el consecuente de la regla (**pred**).

Aunque los arboles de decisión pueden también producir un conjunto de reglas, los métodos de inducción de reglas son diferentes ya que:

- a. Las reglas son independientes y no tienen por qué formar un árbol.
- b. Las reglas generadas pueden no cubrir todas las situaciones posibles.
- c. Las reglas pueden entrar en conflicto en sus predicciones; en este caso; es necesario elegir que regla se debe seguir. Un método para resolver los conflictos consiste en asignar un valor de confianza a las reglas y usar la que tenga mayor confianza.

Las condiciones en el antecedente de las reglas pueden ser comparaciones entre un atributo y uno de los valores de su dominio, o bien entre un par de atributos.

Dentro de esta técnica de minería de datos, tenemos, entre otros, el siguiente algoritmo:

- W-OneR⁷: Éste algoritmo utiliza el atributo de error mínimo para la predicción, discretizando atributos numéricos. Como resultado, el algoritmo devuelve una sola regla.

- Modelo probabilístico: el algoritmo más representativo es Naive Bayes, se basa en la regla de Bayes y que ‘ingenuamente’ asume la independencia de los atributos dada la clase. Este método funciona bien con bases de datos reales, sobre todo cuando se combina con otros procedimientos de selección de atributos que sirven para eliminar la redundancia.

La regla de Bayes establece que, si tenemos una hipótesis H sustentada para una evidencia E, entonces:

$$P(H|E) = [p(E|H) \cdot p(H)]/p(E)$$

donde p(A) representa la probabilidad del suceso A, usando la notación p(A|B) para denotar la probabilidad del suceso A condicionada al suceso B.

1.3. Metodología CRISP-DM

Shearer sostiene que “CRISP-DM (*Cross Industry Standard Process for Data Mining*) es una metodología de minería de datos integral y un modelo de proceso que proporciona a cualquier persona, desde principiantes hasta expertos en minería de datos, un plan

⁷ Obtenido de Documentación de RapidMiner- Descripción del operador W-OneR.

completo para llevar a cabo un proyecto de minería de datos. CRISP-DM desglosa el ciclo de vida de un proyecto de minería de datos en seis fases: comprensión de negocios, comprensión de datos, preparación de datos, modelado, evaluación e implementación.” (Shearer, 2000)

En la Imagen 5 se muestra las fases del proceso de minería de datos. Las flechas indican la relación de dependencias entre fases, mientras que el círculo exterior la naturaleza cíclica de la minería de datos.

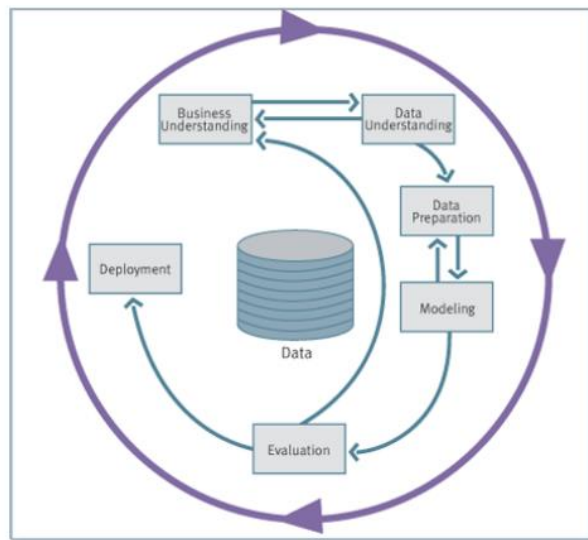


Imagen 5-Fases del proceso de minería de datos. (Shearer, 2000)

La primera fase de la metodología CRISP-DM se denomina entendimiento del negocio, aquí se establecen los objetivos del negocio, se define el problema de minería de datos, para luego elaborar un plan preliminar diseñado para alcanzar los objetivos establecidos. Es de suma importancia conocer el negocio para el cual buscan una solución.

Luego, continuamos hacia la segunda fase (Imagen 6) denominada comprensión de los datos: aquí se realiza la recopilación de los datos inicial. La persona encargada del análisis de estos datos es quien va a familiarizarse con ellos para así poder identificar problemas de calidad de los datos y encontrar (si los hubiera) valores atípicos. Dentro de ésta fase, entonces, tenemos cuatro pasos importantes: la recopilación de datos iniciales, la descripción de los datos, la exploración de datos y la verificación de la calidad de los datos.

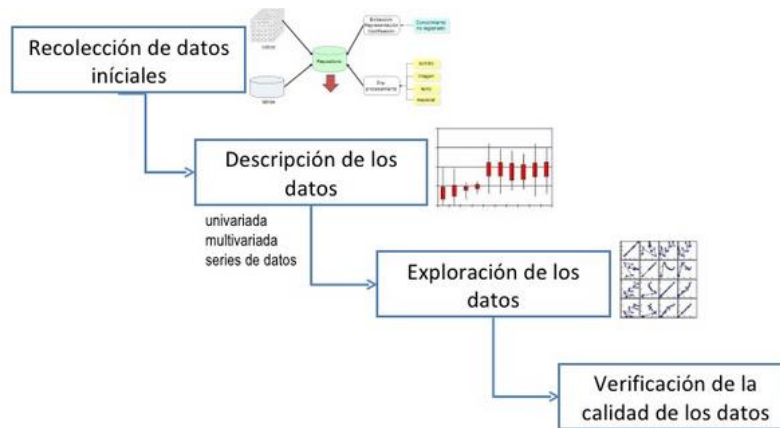


Imagen 6- Fase 2- Minería de datos

Como tercera fase de la metodología CRISP-DM (Imagen 7), encontramos la preparación de los datos, es una de las fases más importantes y también una de las más extensas en cuanto a tiempo, ya que ocupa normalmente el 90% del mismo. En ésta fase lo que se realiza es la evaluación y selección de los datos. Se ignoran aquellos que no aporten al modelo para construir el conjunto de datos final. Además, se realiza una limpieza de los datos para luego prepararlos, por ejemplo, desarrollar atributos completamente nuevos o crear atributos derivados, es decir, atributos nuevos que se construyen a partir de atributos existentes. Otro tipo de atributo derivado son las transformaciones de un solo atributo, que generalmente se realizan para satisfacer las necesidades de las herramientas de modelado. Estas transformaciones pueden ser necesarias para transformar los rangos en campos simbólicos o campos simbólicos a valores numéricos. Las herramientas de modelado o algoritmos a menudo requieren estas transformaciones.

Otra tarea importante desarrollada en esta fase es la integración de los datos. Con ella se mezcla información de distintos registros para crear nuevos. Por ejemplo, si se poseen dos tablas con los mismos atributos y diferentes datos, ambas pueden formar una sola tabla.

Ésta integración también cubre agregaciones, es decir, crear nuevas variables para resumir información de varios registros.

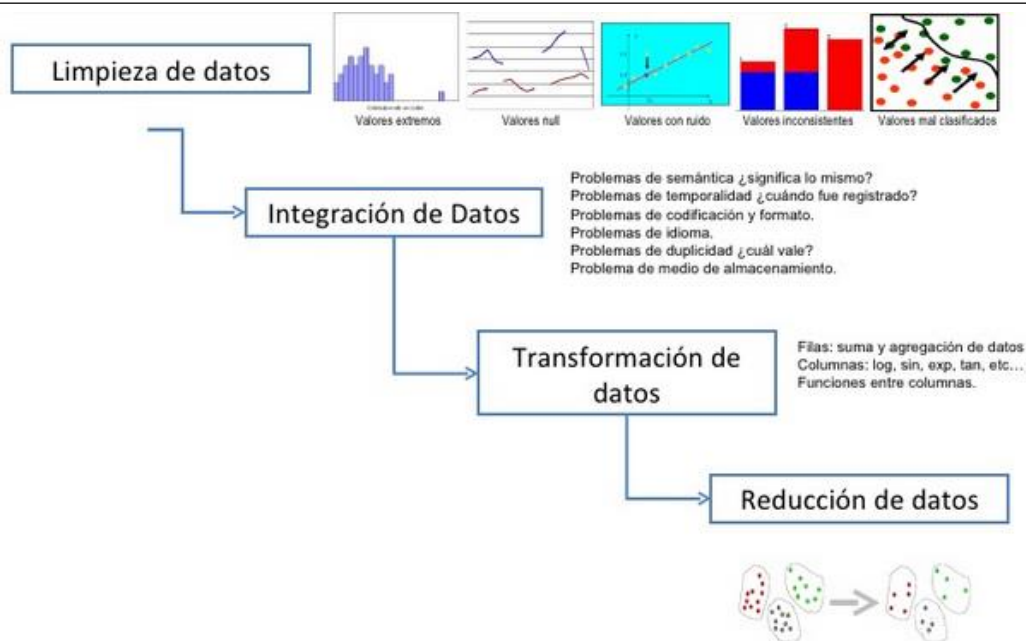


Imagen 7-Fase 3- Minería de datos

La cuarta fase es la construcción del modelo. Existen diversas técnicas de modelado, por lo que se pueden aplicar varias de acuerdo a los objetivos de la minería de datos. Pueden crearse, por ejemplo, árboles de decisión o redes neuronales. Luego de construir el modelo, se procede a realizar pruebas empíricas para determinar la fortaleza del modelo. En otras palabras, se comprueba que tan bien puede predecir el modelo.

Dentro de esta fase se realiza la evaluación del modelo, es decir que, el analista juzga el éxito de la aplicación de técnicas de minería de datos. Éste profesional debe trabajar acompañado del personal del negocio y expertos en dominio para interpretar los resultados de la minería de datos en un contexto empresarial.

La fase cinco de evaluación consiste en evaluar el modelo más a fondo y revisarlo para comprobar que realmente cumpla con los objetivos. El paso anterior evalúa la generalidad y la precisión del modelo. En esta fase se prueban los modelos con nuevos datos reales y también se busca revelar información o sugerencias para futuras direcciones. Además de lo mencionado, se realiza una revisión más exhaustiva del trabajo de extracción de datos para determinar si hay algún factor o tarea importante que de alguna manera se haya pasado por alto. Esta revisión también cubre problemas de control de calidad (por ejemplo, ¿creamos correctamente el modelo? ¿Utilizamos solo los atributos permitidos que están disponibles para la implementación futura?). Al finalizar, se decide si se realizaran o no, más iteraciones.

Por último, la fase de despliegue o implantación. Aquí se presenta el conocimiento obtenido al cliente de manera que éste pueda entenderlo y utilizarlo. Dependiendo de los requisitos, la fase de implementación puede ser tan simple como generar un informe o tan compleja como implementar un proceso de minería de datos repetible en toda la empresa.

1.4. Sistema

Un sistema es un conjunto de elementos interrelacionados que interactúan entre sí y que persiguen un objeto en común. (Yourdon, 1993)

Por otro lado, es necesario definir el concepto de **alarma**. Ésta es entendida como la señal o aviso que advierte sobre la proximidad de un peligro. El aviso de una alarma informará a la comunidad en general o a una entidad específica que deben seguir ciertas instrucciones de emergencia dado que se ha presentado una amenaza. (Pérez Porto & Gardey, Definición de Alarma, 2018)

El término “alarma” es confundido generalmente por el denominado “alerta”. Éste último hace referencia a una situación de vigilancia o atención. (Pérez Porto & Gardey, 2018)

En este trabajo de investigación, utilizamos el término alarma, ya que el sistema contiene una señal o aviso que advierte situaciones de riesgo escolar y no de alerta ya que no se produce ninguna vigilancia como lo manifiesta la definición de la misma.

Por consiguiente, definimos a continuación un **sistema de alarma**. Se entiende a éste como: “un elemento de seguridad pasiva. Esto significa que no evitan una situación anormal, pero si son capaces de advertir de ella, cumpliendo así, una función disuasoria frente a posibles problemas”. (García, 2013)

1.5. Rendimiento académico.

En referencia al ámbito educativo, Vélez define al rendimiento académico como la suma de diferentes y complejos factores que actúan en la persona que aprende, y ha sido definido con un valor atribuido al logro del estudiante en las tareas académicas. Se mide mediante las calificaciones obtenidas, con una valoración cuantitativa, cuyos resultados muestran las materias ganadas o perdidas, la deserción y el grado de éxito académico. (Pérez Luño, Jerónimo, & Sánchez Vasquez, 2000), (Vélez Van & Roa, 2005)

En otras palabras, el rendimiento académico es una medida de las capacidades del alumno, que expresa lo que éste ha aprendido a lo largo del proceso formativo. También supone la capacidad del alumno para responder a los estímulos educativos. En este sentido, el rendimiento académico está vinculado a la aptitud.

Existen distintos factores que inciden en el rendimiento académico. Desde la dificultad propia de algunas asignaturas o cuestiones que están directamente relacionadas al factor psicológico, como la poca motivación, el desinterés o las distracciones, que dificultan la comprensión de los conocimientos impartidos por el docente y termina afectando al rendimiento académico a la hora de las evaluaciones. (Pérez Porto & Gardey, 2018)

Castejón y Pérez describen algunos de los aspectos determinantes influyentes en el rendimiento académico, entre ellos: componentes internos y externos. Pueden ser de orden social, emocional y cognitivo; que se clasifican en tres categorías, y a su vez éstas se clasifican en subcategorías. (Castejón & Pérez , 1998)

La primera de ellas es la denominada *determinantes personales*: aquí se incluyen factores de índole personal, cuyas interrelaciones se pueden producir en función de variables subjetivas, sociales e institucionales. Dentro de los determinantes personales tenemos: la autoevaluación de la propia capacidad del individuo para cumplir una determinada tarea, la percepción sobre su capacidad y habilidades intelectuales. Está relacionada con la influencia ejercida en el entorno familiar e incide en distintas variables que se asocian con el éxito académico tales como: la persistencia, el deseo de éxito, expectativas académicas del individuo y la motivación. El afecto de los padres hacia el estudiante se asocia con el establecimiento de una alta competencia académica percibida y con la motivación hacia el cumplimiento académico. (Pellegrina García & Casanova, 2002)

Otro determinante que influye es *la motivación*, considerada como el impulso que conduce a una persona a elegir y realizar una acción. Existen conceptos relacionados con la motivación, como ser: el vigor, la dedicación y absorción. La dedicación conlleva una alta implicación en las tareas, por lo que experimenta entusiasmo e inspiración; la absorción ocurre cuando se experimenta un alto nivel de concentración y el vigor se caracteriza por altos niveles de energía y resistencia mental, mientras se trabaja, el deseo de invertir esfuerzo en el trabajo que se está realizando incluso cuando aparecen dificultades en el camino. (Salanova, Cifre, Grau, & Martinez, 2005)

Por otro lado, tenemos los factores externos al estudiante, cuya interacción con los determinantes personales da como resultado un estado de motivación. Dentro de los factores externos con los que el alumno puede interactuar, se encuentran aspectos como el compañerismo, la formación docente, condiciones económicas, entre otras; asociados al resultado académico.

Existen casos de estudiantes que por distintas razones se encuentran en estados de agotamiento, desinterés y falta de proyección con sus estudios. La motivación y compromiso de los estudiantes con el logro académico son fundamentales en sus resultados. (Pérez Luño, Jerónimo, & Sánchez Vasquez, 2000)

Por otro lado, tenemos *la asistencia a clases*, la cual se refiere a la presencia del alumno en las lecciones. En una investigación realizada por Perez, Ramon, y Sanchez, se encontró que la motivación está asociada a la asistencia a clases, y que la ausencia se relaciona con problemas de repetición y abandono a los estudios. Cuanta mayor asistencia, mejor calificación; la asistencia es una de las variables significativas en el rendimiento académico del alumno. (Pérez Luño, Jerónimo, & Sánchez Vasquez, 2000)

Como segunda clasificación tenemos a *los determinantes sociales* que son factores de índole social que interactúan con la vida académica del estudiante. Aquí encontramos por un lado el *entorno familiar*; entendido como el conjunto de interacciones propias de la convivencia familiar. Las influencias de los padres o del adulto responsable, influye en el rendimiento escolar del alumno. Cuando un estudiante se encuentra en un ambiente

autoritario e indiferente de parte de sus padres, incide de manera negativa en el desempeño del alumno.

En una investigación realizada por la UTN se encontraron factores influyentes en el rendimiento académico relacionados con los aspectos sociales nombrados anteriormente. Luego de aplicar técnicas de recolección de datos, como ser: revisión de documentos, se identificaron factores como ser: *nivel de estudio de la madre*. Se observó que el 22% de los alumnos que obtuvieron un buen rendimiento académico tienen madres con estudios de grado y postgrados. También se determinó como otro factor influyente a la *cantidad de horas semanales* que los alumnos dedican al estudio. En la muestra seleccionada de 50 alumnos cursantes de la cátedra “Estructura de Datos y Algoritmos”, se observó que el 56% de quienes dedican más de 15 horas semanales al estudio, poseen buen rendimiento académico. Otro factor identificado durante la investigación es la *importancia que los alumnos otorgan al estudio*. Se observó que el 80% de quienes han tenido rendimiento académico bueno han otorgado más importancia al estudio que a la diversión. Además, el 11% de quienes han tenido un desempeño aceptable han otorgado más importancia al estudio que al *trabajo*. Esto indica una relación entre el éxito académico y la importancia otorgada a los estudios y al trabajo.

En estudios realizados por Carrión establece que las *condiciones demográficas*, es decir zona geográfica en la que vive el estudiante en época lectiva son factores que eventualmente se relacionan con el rendimiento académico en forma positiva o negativa.

Por último, la categoría de *determinantes institucionales*. Según Carrión son los componentes no personales que intervienen en el proceso educativo que influyen en el rendimiento académico. Dentro de este destacamos la dificultad de las materias y condiciones institucionales. (Carrión Perez, 2002)

La *complejidad de los estudios* se refiere a la dificultad de algunas materias de las áreas académicas. Según estudios realizados, la destacaron como una variable importante en el rendimiento académico. (Salanova, Cifre, Grau, & Martinez, 2005)

Las *condiciones institucionales* como ser el plan de estudios, servicios: como ser el apoyo que brinda la institución: sistemas de becas, servicio de préstamos de libros, asistencia médica, apoyo psicológico, entre otros; y condiciones de las aulas también pueden afectar el rendimiento académico como así también facilitarlos.

1.6. Riesgo académico

Juan Vega define al riesgo académico como la posibilidad de que el alumno sea dado de baja académicamente, es decir, se está en riesgo de no poder avanzar en la institución educativa debido a las calificaciones obtenidas, al desempeño académico. (Vega, 2015)

En otras palabras, el riesgo académico surge cuando se poseen calificaciones no superadoras durante el ciclo lectivo, lo cual, en el caso del nivel medio, el alumno tiende a repetir de curso.

El riesgo académico abarca tanto la probabilidad de abandonar la escuela como de tener un nivel de desempeño por debajo de una línea de base previamente establecida para ese país. (Muelle, 2016)

Por esta razón, es importante hacer un seguimiento del rendimiento académico del cada alumno para evitar que su trayectoria escolar se vea afectada.

1.7. Repitencia escolar

Se define a la repitencia escolar como la situación que se produce cuando el alumno no es promovido al curso siguiente, debido a sus bajas calificaciones; por lo que debe repetir aquel que estaba cursando nuevamente. (Pérez Porto & Gardey, Definición de repitencia escolar, 2018)

Como es un problema cuya ocurrencia depende del bajo rendimiento escolar, es de suma importancia comprenderlo.

1.8. Investigaciones previas

Existen investigaciones relacionadas a la temática del presente trabajo de investigación, por ello se describirán aquellas que fueron más relevantes.

En una investigación realizada por Orea, Vargas y Alonso acerca de las causas por las que los alumnos abandonan la facultad, se logró identificar a aquellos alumnos con posibilidad de deserción por medio de la aplicación de la minería de datos.

Se obtuvieron bases de datos de la universidad tecnológica de México, de 723 alumnos que fueron inscriptos y posteriormente dados de baja. Esta información recolectada les permitió realizar el análisis, limpieza y selección de datos, con el fin de obtener atributos relevantes para realizar la correspondiente vista minable. Ésta vista, contuvo los siguientes campos: sexo, edad, tipo de título secundario obtenido, promedio final del secundario, apoyo económico, inglés, educación del padre, educación de la madre, ingreso familiar, tamaño de la familia, trabaja, horas de trabajo. (Orea, Vargas, & ALonso Garcia, 2009)

Las técnicas de minería de datos utilizadas en este trabajo fueron: k vecinos más cercanos y árboles de decisión. La precisión obtenida del algoritmo árbol de decisión fue de un 98,95%, mientras que en k vecinos más cercanos fueron las siguientes: con k=1 fue de 62,51%, con k=10 fue de un 67,08%, con k=50 de 67,77% y con k=100 de 67,63%. Por lo tanto, se eligió el algoritmo de árbol de decisión ya que presenta una confiabilidad mayor, en comparación con los demás resultados.

En otra investigación, realizada en Colombia se muestra la construcción de un modelo predictivo para determinar la probabilidad de deserción de los estudiantes de la

universidad Simón Bolívar. Aquí se obtuvieron datos de los alumnos registrados en los sistemas de información de la universidad, como ser datos personales, socio-económicos y los derivados del desempeño académico de los estudiantes durante su permanencia en la institución; entre ellos: edad, ciudad de procedencia, estrato, sexo, jornada, promedio, materias perdidas, ocupación, materias cursadas, nivel de estudios del padre, nivel de estudio de la madre, e ingresos, teniendo así, un total de 201 instancias.

Para obtener un modelo predictivo adecuado utilizaron la técnica de árbol de decisión C4.5 e ID3. Al comparar la precisión de ambos, se eligió el árbol ID3 debido a que poseía la mayor precisión al momento de predecir qué alumnos podían desertar en dicha universidad. (Amaya Torrado, Avendaño Barrientos, & Heredia Vizcaíno, 2012)

Por último, un estudio realizado en Argentina en la universidad de Corrientes determinó que el riesgo académico de los alumnos de primer año, de la cátedra *Estructura de datos y algoritmos* de la carrera Ingeniería en sistemas de la universidad, estaba relacionado, entre otros, con factores como: nivel educativo de la madre, título secundario, obra social, sexo y la carrera elegida. (La Red Martínez, Karanik, Giovanni, & Scappini, 2016)

En ésta investigación obtuvieron mediante la aplicación de la minería de datos, un modelo basado en técnicas de regresión, cuya precisión fue la más alta en comparación con otras técnicas.

Como se puede observar en las investigaciones nombradas, se analizaron algunos problemas relacionados a la educación con técnicas de minería de datos, y a pesar de que fueron realizadas en el nivel universitario se observa una aproximación al presente trabajo.

La repitencia escolar, el bajo rendimiento académico y otros problemas en el nivel medio han sido tratados ampliamente, sin embargo, se desconoce que se haya realizado con técnicas o metodologías que presenta este trabajo.

Capítulo 2: Descripción del problema.

La Escuela de Comercio N° 5091 Prof. Julio Cortes, ubicada en la ciudad de Oran, provincia de Salta; no es ajena a las problemáticas del bajo rendimiento y repitencia escolar debido a la cantidad de alumnos desaprobados y repitentes.

Acorde a los datos recolectados, se observa un incremento de alumnos desaprobados entre el periodo lectivo 2016 y el periodo lectivo 2017 en el tercer año de estudio. Entendiéndose por desaprobados a los alumnos que obtienen calificaciones inferiores a 6, según el régimen de evaluación, calificación y promoción del año 1992 del colegio secundario de Argentina⁸.

Se conoce, acorde a investigaciones previas, que el rendimiento de los alumnos de debe a múltiples factores (Magaña Hernández, 2002), pero en este colegio los actores educativos desconocen los mismos y además no cuentan con alguna herramienta que pueda dar solución a la problemática existente, como ser el conocimiento anticipado de la situación académica de los alumnos como para establecer o planificar estrategias que prevengan el bajo rendimiento académico y así evitar la repitencia.

Así mismo, los preceptores también desconocen los factores que podrían influenciar en el bajo rendimiento de los alumnos a cargo. Éstos solo cuentan con registros de notas y asistencias, pero no poseen herramientas que sean capaz de prevenir el bajo rendimiento de los estudiantes.

Además, los directores se encuentran imposibilitados a realizar gestiones a tal problemática de manera anticipada debido a que reciben informes de sus docentes una vez que el alumno adquiere una nota que determina la situación del mismo, lo que les impide gestionar anticipadamente acciones para prevenir situación de bajo rendimiento.

Por otro lado, cuando un padre o tutor de un alumno visita el establecimiento para obtener información académica del mismo, éste solo obtiene la calificación final del trimestre de cada espacio curricular. En el caso de que el alumno este desaprobado solo se conoce la nota menor a 6, no así el factor que la podría haber ocasionado.

El establecimiento posee y utiliza herramientas ofimáticas (Word y Excel) las cuales solo posibilitan el registro de los alumnos por curso con respectivos datos como ser: nombre completo, DNI, tutor y dirección y la situación escolar relacionada a notas y asistencias. Y no más datos relacionados al ámbito social, laboral y personal de los estudiantes.

En base a lo manifestado e investigado, se propone un sistema informático denominado sistema de alarma basado en la minería de datos, el cual tiene por objetivo identificar con anticipación situaciones de riesgo académico del alumnado que concurre a este

⁸ Régimen de evaluación, calificación y promoción:

<http://www.bnm.me.gov.ar/giga1/documentos/EL003757.pdf>

establecimiento atendiendo de esta manera a las demandas de todos los actores educativos y a la calidad educativa solicitada por el ministerio de educación.

2.1. Objetivo General

Desarrollar un sistema de alarma para la detección del riesgo académico basado en un modelo predictivo de minería de datos en un colegio secundario.

2.1.2. Objetivos específicos

- Aplicar técnicas de recolección de datos.
- Tabular datos obtenidos.
- Aplicar la minería de datos para obtener un modelo predictivo a fin de identificar el riesgo escolar y los factores que influyen en el bajo rendimiento académico.
- Visualizar resultados.
- Diseñar el prototipo del sistema de alarma.
- Incorporar el modelo predictivo al prototipo del sistema.

2.2. Alcance

El sistema de alarma desarrollado dispone de una alarma de riesgo escolar según calificaciones, inasistencias, y otros factores, la cual permite anticiparse a situaciones de riesgo educativo. La alarma permite advertir situaciones de riesgo escolar conocidas como desconocidas, entre las conocidas se encuentran el riesgo por inasistencias y calificaciones, y entre lo desconocido, la información que aporta el modelo.

Este sistema, no controla asistencias de los docentes y directivos, y no incluye funciones administrativas. Además, no necesita una conexión a internet para su correcto funcionamiento.

También permite tener un control de acceso de los usuarios debido a que la información que se administra en el sistema es de carácter sensible.

Además, el sistema de alarma tiene un alcance informativo para toda la comunidad educativa porque provee información personalizada de cada estudiante que concurre a la institución. Los actores institucionales: preceptores y directivos, pueden acceder al sistema en el momento en que se lo requiera y proveer dicha información a los padres y/o tutores.

Consecuentemente, el sistema informático basado en la minería de datos provee beneficios a la institución educativa, ya que a través de ella se posee información rápida e importante de la situación académica del alumnado. Sin embargo, posee una desventaja tecnológica en el momento de implementar dicho sistema debido a la antigüedad de las computadoras que posee actualmente la institución.

El establecimiento solo cedió los datos (notas e inasistencias) del primer trimestre de los alumnos pertenecientes al 3° año 1° división del turno tarde, por lo cual, las encuestas y el análisis de este trabajo se realiza sobre éstos alumnos. Aunque la cantidad de registros cedidos es escasa, no es una limitación para el entrenamiento de un modelo predictivo.

La propuesta está pensada para que a medida que se carguen datos de otros cursos y/o divisiones, el modelo de minería, se re-entrene cada determinado tiempo. Esto significa, que no será necesario realizar el proceso de minería nuevamente.

2.3. Alternativas tecnológicas

Para concretar y realizar el sistema de alarma y el proceso de minería de datos, se analizaron distintas alternativas tecnológicas mencionadas a continuación:

- Para minería de datos:
 - RapidMiner Studio: es una herramienta de minería de datos, que implementa técnicas de pre-procesamiento de datos, modelación predictiva, modelación descriptiva, métodos de prueba de modelos, visualización de datos y transformación de datos⁹.
 - Weka: es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Contiene herramientas para la preparación de datos, clasificación, regresión, agrupamiento, minería de reglas de asociación y visualización. Weka es un software de código abierto emitido bajo licencias publicas general GNU¹⁰.

- Para el desarrollo del prototipo:
 - C#: el lenguaje C# es un lenguaje de programación orientado a objetos desarrollado y estandarizado por Microsoft. Permite a los desarrolladores crear una gran variedad de aplicaciones seguras y sólidas que se ejecutan en .NET Framework. Visual C# proporciona un editor de código avanzado, prácticos diseñadores de interfaz de usuario, un depurador

⁹ RapidMiner Studio: <https://rapidminer.com/>

¹⁰ Página Oficial Weka: <https://www.cs.waikato.ac.nz/ml/weka/>

integrado y muchas otras herramientas que facilitan el desarrollo de aplicaciones basadas en el lenguaje C# y .NET Framework¹¹.

- Java: es un lenguaje de programación orientado a objetos, creado para crear aplicaciones robustas y fiables. Los programadores pueden escribir el código una sola vez y ejecutarlo en cualquier dispositivo. Esto es posible gracias a que Java cuenta con una Java Virtual Machine (JVM) que brinda portabilidad al lenguaje. JDK posee un compilador que toma el código Java y valida la sintaxis, si el compilador encuentra errores mostrará un mensaje, sino, crea un archivo byte.class, el cual será ejecutado por JVM¹².
- Javascript: es un lenguaje de programación interpretado, utilizado principalmente en su forma del lado del cliente, implementado como parte de un navegador web permitiendo mejoras en la interfaz de usuario y páginas web dinámicas¹³.

➤ Bases de datos relacionales:

- MySQL: es un sistema de gestión de base de datos cuya finalidad es el desarrollo web. MySQL trabaja con bases de datos relacionales, es decir utiliza tablas múltiples que se interconectan entre sí para almacenar la información y organizarla correctamente¹⁴.
- SQL Server: SQL Server es un sistema de gestión de bases de datos relacional (RDBMS) usado para manipular y recuperar datos, crear tablas y definir relaciones entre ellas¹⁵.

La elección final de las herramientas se detalla en la sección 3.1.

¹¹ Página Oficial de Microsoft <https://docs.microsoft.com/es-es/dotnet/csharp/getting-started/introduction-to-the-csharp-language-and-the-net-framework>

¹² Java: <https://devcode.la/blog/que-es-java/>

¹³ Javascript: <https://es.wikipedia.org/wiki/JavaScript>

¹⁴ MySQL: <https://neoattack.com/neowiki/mysql/>

¹⁵ ¿Qué es SQL Server Management Studio? <https://searchdatacenter.techtarget.com/es/definicion/SQL-Server>

Capítulo 3: Solución propuesta

Se propone como solución a la problemática planteada la aplicación de la minería de datos para obtener un modelo predictivo que permita identificar de manera anticipada situaciones de riesgo académico.

De esta manera será posible identificar, aquellos factores que posiblemente influyen en el rendimiento del alumno.

El modelo obtenido será incorporado a un sistema de alarma para facilitar la interacción con los actores educativos.

Este sistema, utiliza el término alarma para advertir situaciones conocidas y desconocidas. Entre las conocidas se encuentran las inasistencias y calificaciones, y desde lo desconocido: la información que aporta el modelo.

3.1. Herramientas utilizadas

Las herramientas utilizadas en el trabajo son:

- Para el proceso de minería de datos:
 - RapidMiner Studio. La elección del mismo se debe a su fácil e intuitivo uso.
- Para el diseño del prototipo:
 - Lenguaje C#. La elección del este, se debe a la experiencia en desarrollo con dicho lenguaje.
- Para las bases de datos:
 - Su elección se debe a la fácil integración con el lenguaje C# y la experiencia en el uso de la misma.

3.2. Desarrollo de la solución

Se desarrolla a continuación la solución propuesta en base a la metodología CRISP-DM detallada en la sección 1.3. La elección de la misma se debe a que es una de las más específicas, detalladas, y adecuadas para la ejecución del proyecto.

- **Fase 1: Entendimiento del negocio**

En nuestro caso, la expresión “negocio” hace referencia a la institución educativa.

Primeramente, se solicitó autorización mediante una nota dirigida al director del establecimiento, para realizar el trabajo de investigación en la institución. Luego de obtener dicha autorización, se realizaron entrevistas al personal del establecimiento con el fin de identificar la problemática existente, forma de trabajo, tecnologías disponibles, cantidad de alumnos que concurren a la misma, comportamientos de los alumnos, horarios de clases, tipos de sanciones, cantidad de turnos, cantidad de cursos y divisiones.

Con el objeto de obtener más información personal sobre los alumnos, la cual no se encuentra en algún registro perteneciente a la institución, se procedió a realizar encuestas (no anónimas) a los alumnos del tercer año primera división pertenecientes al turno tarde de la institución.

A continuación, se muestra el modelo de encuesta realizado:

ENCUESTA.

Por favor marcar con X la respuesta correcta.

¡¡Muchas gracias por su colaboración!!

Nombre y Apellido: _____

Institución: _____

Fecha: __/__/__

1. ¿Qué edad tienes?

- Menos de 16
- 16
- 17
- Más de 17

2. ¿Con quién vives?

- Solo
- Padre
- Madre
- Otros familiares

3. ¿Cuántos hermanos tienes?

- Ninguno
- 1 o 2
- Más de 2
- Más de 5

4. ¿Eres el mayor de tus hermanos?

SI

NO

Imagen 8-Modelo de encuesta (Parte 1)

5. Nacionalidad:
- Argentina
 - Otra
6. ¿En qué zona vivís?
- Zona rural
 - Zona urbana
7. ¿Trabajas?
- SI NO
8. ¿Realizas alguna actividad extraescolar?
- Deporte
 - Idioma
 - Otros
9. ¿Qué materia/s te costaron más en 3er año?

Imagen 9-Modelo de encuesta (Parte 2)

- Administración
 - Arte
 - Biología
 - Economía
 - Educación Física
 - Geografía
 - Historia
 - Lengua Extranjera
 - Lengua y Literatura
 - Matemática
 - Sistema de información contable
10. ¿Repetiste 3er año?
- SI NO
11. ¿Viniste de pase de otro establecimiento en 3er año?
- SI NO
12. ¿Cuánto tiempo le dedicas a los estudios?
- Menos de 1 hora por día
 - Entre 1 y 2 horas por día
 - Entre 3 y 4 horas por día
 - Más de 3 horas por día
13. ¿Sos beneficiario de algún plan o beca?
- SI NO

Imagen 10-Modelo de encuesta (Parte 3)

14. Distancia de la casa a la escuela (aproximadamente):

- Menos de 3 cuadras
- Entre 3 y 10 cuadras
- Más de 10 cuadras

15. ¿Cómo se traslada a la escuela?

- Caminando
- Colectivo
- Otro

16. Tu tutor/a es:

- Madre
- Padre
- Abuela/o
- Otro

17. ¿Tu tutor/a trabaja?

SI

NO

¡¡MUCHAS GRACIAS!!

Imagen 11-Modelo de Encuesta (parte 4)

La institución cedió los registros de calificaciones y asistencias de los alumnos. La Imagen 12 corresponde al registro de calificaciones de un alumno en particular y la Imagen 13 a la planilla de asistencia del mes de mayo.

La Autoridad del Establecimiento educativo: ESCUELA DE COMERCIO N° 5091 "Prof. JULIO CORTES" C.U.E.N° 6600647 00 de la Ciudad de San Ramón de la Nueva Orán - Dpto. ORAN - Pcia. de SALTA; hace constar que el/la alumno/a:

DNI: [REDACTED] Fecha de Baja: Estado del Alumno: 12/218

Obser 2017: CSO: 0 Libro 0 Folio 0

Materia: 1° Trim 2° Trim 3° Trim Cal.Final I.E.Dic. I.E.Feb. Cal.Def. Acta: Fecha Examen: Obser1:

Condición: **Regular**
 AÑO: 2017 Curso: 3°1° Turno: Tarde Modalidad: Bachiller en Economía y Administración
 CURSO: 31T
 Opcion: Obligatoria

| | 9 | 10 | 8 | 9 | 0 | 0 | 9 | |
|---------------------------------|----|----|----|----|---|---|----|----------|
| Administración | 8 | 10 | 10 | 10 | 0 | 0 | 10 | Aprobado |
| Arte | 10 | 9 | 10 | 10 | 0 | 0 | 10 | Aprobado |
| Biología | 9 | 7 | 8 | 8 | 0 | 0 | 8 | Aprobado |
| Economía | 8 | 10 | 10 | 9 | 0 | 0 | 9 | Aprobado |
| Educación Física | 8 | 8 | 8 | 8 | 0 | 0 | 8 | Aprobado |
| Geografía | 7 | 9 | 8 | 8 | 0 | 0 | 8 | Aprobado |
| Historia | 9 | 7 | 6 | 6 | 0 | 0 | 6 | Aprobado |
| Lengua Extranjera | 8 | 6 | 9 | 8 | 0 | 0 | 8 | Aprobado |
| Lengua y Literatura | 9 | 9 | 10 | 9 | 0 | 0 | 9 | Aprobado |
| Matemática | 8 | 10 | 10 | 10 | 0 | 0 | 10 | Aprobado |
| Sistema de Información Contable | | | | | | | | |

Observac_Institucionales: PromGral: lunes, 26 de marzo de 2018

Observaciones: 8,636

ObserCur:

IMPORTANTE: MATERIAS PENDIENTES DE APROBACION SUJETAS A VERIFICACION ADMINISTRATIVA




Imagen 12-Registro de calificaciones

Escuela de Comercio Nº 5091 "Profesor Julio Cortes"
Control de Asistencia del mes de **Mayo** de 2017

| 2017 | Tarde | Bachiller en Economía y Administración | DNI | Observar 2017 | Días del mes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Anterior | | Actual | | | |
|------|-------|--|-----|---------------|--------------|--------|-----------|--------|---------|--------|---------|-------|--------|-----------|--------|---------|--------|---------|-------|--------|-----------|--------|---------|--------|---------|-------|--------|-----------|--------|---------|--------|---------|-------|-------|-------|----------|------|--------|--------|------|-------|
| | | | | | LUNES | MARTES | MIERCOLES | JUEVES | VIERNES | SABADO | DOMINGO | LUNES | MARTES | MIERCOLES | JUEVES | VIERNES | SABADO | DOMINGO | LUNES | MARTES | MIERCOLES | JUEVES | VIERNES | SABADO | DOMINGO | LUNES | MARTES | MIERCOLES | JUEVES | VIERNES | SABADO | DOMINGO | Total | Total | Total | Total | | | | | |
| 1 | | | | | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 2/4 | 135 | 25 | 5 | | |
| 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 24/4 | 18 | 2 | 26,5 | | |
| 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 19 | 1 | 2 | | |
| 4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 17 | 3 | 3 | | |
| 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 20 | - | - | | |
| 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 17/4 | 17 | 3 | 4,5 | |
| 7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 9 | 13,5 | 2,5 | 11,5 | |
| 8 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 5 | 18 | 2 | 7 | |
| 9 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 3/4 | 19 | 1 | 4,5 | |
| 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 18,5 | 1,5 | 2,5 | |
| 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 20 | - | - | |
| 12 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1/2 | 20 | - | 1/2 | |
| 13 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 20 | - | - | |
| 14 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 3/4 | 15,5 | 4,5 | 8 | |
| 15 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 3 | 19 | 1 | 4 |
| 16 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1/2 | 20 | - | 1/2 |
| 17 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 20 | - | - | |
| 18 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 4 | 14,5 | 5,5 | 9,5 | |
| 19 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 4 1/2 | 18,5 | 7,5 | 12,5 |
| 20 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 5 | 18,5 | 7,5 | 12,5 |
| 21 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 2 1/2 | 20 | - | 1 |
| 22 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 20 | - | 1 |
| 23 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 19 | 1 | 1 |
| 24 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 20 | - | - |
| 25 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 5 | 15 | 2 | 3 2,8 |
| 26 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 5 1/2 | 16 | 4 | 9,5 |
| 27 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 4 | 15,5 | 4,5 | 8,5 |
| 28 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 2 | 18 | 2 | 4 |
| 29 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 1/2 | 17 | 3 | 4,5 |
| 30 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 3 1/2 | 18 | 2 | 9,5 |
| 31 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 19 1/2 | 1/2 | 1,5 |
| 32 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 20 | - | - |
| 33 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 3 | 17 | 3 | 6 |
| 34 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 14,5 | 1/2 | 1/2 |
| 35 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 20 | - | - | |

Imagen 13-Registro de asistencias

▪ **Fase 2: Entendimiento de datos:**

En esta fase se realizó un análisis de los datos recolectados para verificar la calidad de los mismos.

De los 32 alumnos de 3er año 1ra división que figuraban en los registros de calificaciones, se completaron 30 encuestas (en formato papel) debido a que 2 alumnos estuvieron ausentes cuando se las realizó.

Además, se observó que los registros de notas, asistencias y encuestas estaban completos.

Por otro lado, no se encontraron valores atípicos en las encuestas ya que, los alumnos debían seleccionar una o más opciones en cada pregunta. Tampoco se encontraron valores atípicos en los registros de la institución ya que estos son completados rigurosamente por los preceptores.

Por ello, posterior al análisis y familiarización con los datos obtenidos, se procedió a realizar la tabulación e integración de los mismos en una planilla Excel. Como

las encuestas no eran anónimas, se logró integrar los datos (encuestas y registros de calificaciones) con los nombres de los alumnos.

Dicha planilla se muestra en la Imagen 14, Imagen 15 e Imagen 16.

| Nombre | Edad | Administración | Arte | Biología | Economía | Educación | Geografía | Historia | Lengua I | Lengua II | Matemática | Sistemas | Previa |
|------------|------|----------------|------|----------|----------|-----------|-----------|----------|----------|-----------|------------|----------|--------|
| Ruben | 17 | 4 | 9 | 8 | 4 | 5 | 7 | 6 | 3 | 3 | 4 | 1 | 1 |
| Gaston | 15 | 4 | 9 | 10 | 6 | 8 | 5 | 6 | 8 | 4 | 6 | 1 | 2 |
| Francisco | 15 | 9 | 6 | 10 | 6 | 9 | 7 | 6 | 7 | 7 | 7 | 6 | 0 |
| Eric | 16 | 9 | 8 | 10 | 9 | 8 | 8 | 7 | 9 | 8 | 9 | 8 | 0 |
| Victor | 15 | 4 | 6 | 10 | 8 | 8 | 5 | 7 | 4 | 5 | 8 | 6 | 1 |
| Franco | 16 | 6 | 4 | 7 | 6 | 8 | 6 | 6 | 9 | 5 | 5 | 1 | 0 |
| Diego | 15 | 6 | 8 | 10 | 6 | 5 | 5 | 7 | 3 | 6 | 6 | 1 | 2 |
| Juan | 15 | 4 | 7 | 9 | 7 | 5 | 6 | 6 | 5 | 3 | 6 | 1 | 0 |
| Juan Gabr | 16 | 8 | 8 | 10 | 7 | 6 | 8 | 8 | 9 | 6 | 7 | 6 | 0 |
| Agustin | 15 | 6 | 4 | 10 | 6 | 5 | 7 | 6 | 4 | 6 | 8 | 1 | 2 |
| Alejandro | 15 | 6 | 3 | 9 | 6 | 6 | 6 | 6 | 7 | 5 | 6 | 6 | 2 |
| Walter | 16 | 4 | 9 | 7 | 4 | 8 | 7 | 5 | 8 | 6 | 8 | 6 | 1 |
| Pablo | 16 | 6 | 7 | 10 | 8 | 6 | 6 | 6 | 7 | 6 | 7 | 6 | 0 |
| Candela | 15 | 4 | 8 | 7 | 4 | 6 | 7 | 6 | 9 | 6 | 4 | 1 | 1 |
| Melina | 17 | 8 | 8 | 6 | 9 | 6 | 8 | 5 | 3 | 6 | 2 | 1 | 1 |
| Brisa | 15 | 10 | 7 | 10 | 9 | 7 | 8 | 7 | 9 | 5 | 9 | 6 | 0 |
| Romina | 15 | 6 | 9 | 10 | 8 | 7 | 7 | 9 | 8 | 6 | 9 | 6 | 0 |
| Pamela | 15 | 7 | 4 | 8 | 4 | 5 | 4 | 6 | 3 | 5 | 1 | 1 | 2 |
| Cintia | 16 | 6 | 8 | 9 | 6 | 8 | 7 | 6 | 4 | 4 | 9 | 1 | 1 |
| Magali | 16 | 10 | 10 | 10 | 9 | 8 | 9 | 10 | 9 | 9 | 10 | 10 | 0 |
| Rocio Viln | 16 | 4 | 3 | 10 | 4 | 2 | 5 | 5 | 2 | 3 | 2 | 1 | 2 |
| Natalia | 17 | 4 | 8 | 9 | 4 | 7 | 7 | 4 | 8 | 3 | 9 | 6 | 1 |

Imagen 14-Excel generado (parte 1)

| M | N | O | P | Q | R | S | T | U | V | W | X |
|---------|--------|-----------------|---------|----------|-------------|---------|---------|-----------|-----------------|-----------|------|
| Sistema | Previa | Convivencia | Cantida | Mayor de | Nacionalida | Zona | Trabaja | Actividad | Dificultad e | Repitente | Pase |
| 1 | 1 | Padre, Madre | 3 | NO | Argentina | Urbana | SI | NO | SIC, Matemática | SI | NO |
| 1 | 2 | Padre | 2 | NO | Argentina | Urbana | NO | Deporte | Matemática | NO | NO |
| 6 | 0 | Padre, Madre | 0 | SI | Argentina | Urbana | NO | Idioma | SIC | NO | NO |
| 8 | 0 | Padre, Madre, H | 1 | SI | Argentina | Urbana | NO | Deporte | Historia | NO | NO |
| 6 | 1 | Otros familiare | 3 | NO | Argentina | Urbana | NO | NO | Administración | NO | NO |
| 1 | 0 | Padre, Madre, H | 1 | SI | Argentina | Urbana | NO | Idioma | SIC, Arte, M | NO | SI |
| 1 | 2 | Padre | 2 | NO | Argentina | Urbana | NO | NO | SIC, Lengua | NO | NO |
| 1 | 0 | Madre, Herman | 3 | NO | Argentina | Urbana | NO | Otro | SIC, Admini | NO | SI |
| 6 | 0 | Padre, Madre | 4 | NO | Argentina | Urbana | NO | Otro | Matemática | NO | NO |
| 1 | 2 | Padre, Madre, C | 1 | SI | Argentina | Urbana | NO | NO | SIC, Lengua | NO | NO |
| 6 | 2 | Padre, Madre | 1 | NO | Otra | Urbana | NO | Deporte | Administración | NO | NO |
| 6 | 1 | Madre | 2 | NO | Argentina | Urbana | NO | NO | Economía | NO | NO |
| 6 | 0 | Padre, Madre | 0 | SI | Argentina | Urbana | NO | NO | Ninguna | NO | NO |
| 1 | 1 | Padre, Madre, H | 1 | SI | Argentina | Urbana | NO | NO | SIC, Admini | NO | NO |
| 1 | 1 | Madre | 3 | NO | Otra | Rural | NO | NO | SIC, Matem | SI | NO |
| 6 | 0 | Padre, Madre, C | 6 | NO | Argentina | Urbana | NO | NO | Ninguna | NO | NO |
| 6 | 0 | Ausente | Ausente | Ausente | Ausente | Ausente | Ausente | Ausente | Ausente | NO | NO |
| 1 | 2 | Padre, Madre | 7 | NO | Argentina | Urbana | SI | Idioma | SIC, Educaci | NO | NO |
| 1 | 1 | Padre, Madre | 3 | SI | Argentina | Urbana | NO | Otro | Ninguna | SI | NO |
| 10 | 0 | Padre, Madre | 0 | SI | Argentina | Urbana | NO | Idioma, E | Educación F | NO | NO |
| 1 | 2 | Padre, Madre | 6 | NO | Argentina | Urbana | NO | Deporte | Matemática | NO | NO |
| 6 | 1 | Madre | 1 | SI | Argentina | Urbana | SI | NO | Economía, A | SI | NO |

Imagen 15-Excel generado (parte 2)

| e | Pase | Tiempo ded | Plan/Beca | Distancia casa-in | Movilidad | Tutor | Trabaja | Riesgo Acac |
|---|------|---------------|-----------|-------------------|-----------|---------|---------|-------------|
| | NO | Entre 1 y 2 h | NO | Entre 3 y 10 cuad | Caminando | Padre | SI | Alto |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Caminando | Madre | SI | Alto |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Otro | Padre | SI | Bajo |
| | NO | Entre 3 y 4 h | SI | Entre 3 y 10 cuad | Caminando | Padre | SI | Bajo |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Colectivo | Abuelo | NO | Alto |
| | SI | Entre 3 y 4 h | NO | Entre 3 y 10 cuad | Caminando | Padre | SI | Alto |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Colectivo | Padre | SI | Alto |
| | SI | Entre 1 y 2 h | SI | Mas de 10 cuadra | Colectivo | Madre | SI | Alto |
| | NO | Entre 1 y 2 h | SI | Menos de 3 cuad | Otro | Padre | SI | Bajo |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Colectivo | Padre | SI | Alto |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Otro | Madre | NO | Alto |
| | NO | Entre 3 y 4 h | SI | Mas de 10 cuadra | Caminando | Madre | SI | Alto |
| | NO | Menos de 1 | NO | Mas de 10 cuadra | Caminando | Padre | SI | Bajo |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Colectivo | Padre | SI | Alto |
| | NO | Mas de 3 hs | NO | Mas de 10 cuadra | Colectivo | Madre | SI | Alto |
| | NO | Entre 3 y 4 h | SI | Mas de 10 cuadra | Colectivo | Madre | SI | Bajo |
| | NO | Ausente | Ausente | Ausente | Ausente | Ausente | Ausente | Bajo |
| | NO | Menos de 1 | NO | Mas de 10 cuadra | Caminando | Padre | SI | Alto |
| | NO | Mas de 3 hs | NO | Mas de 10 cuadra | Caminando | Padre | SI | Alto |
| | NO | Mas de 3 hs | SI | Menos de 3 cuad | Caminando | Padre | Si | Bajo |
| | NO | Entre 1 y 2 h | NO | Entre 3 y 10 cuad | Caminando | Madre | SI | Alto |
| | NO | Entre 1 y 2 h | NO | Mas de 10 cuadra | Colectivo | Madre | SI | Alto |

Imagen 16-Excel generado (parte 3)

Como se puede observar en la Imagen 15 e Imagen 16, para los alumnos que no pudieron completar la encuesta debido a su ausencia, se completaron los campos faltantes con la palabra “Ausente”.

Luego, se importó este archivo Excel al programa RapidMiner como se puede observar en la Imagen 17 e Imagen 18:

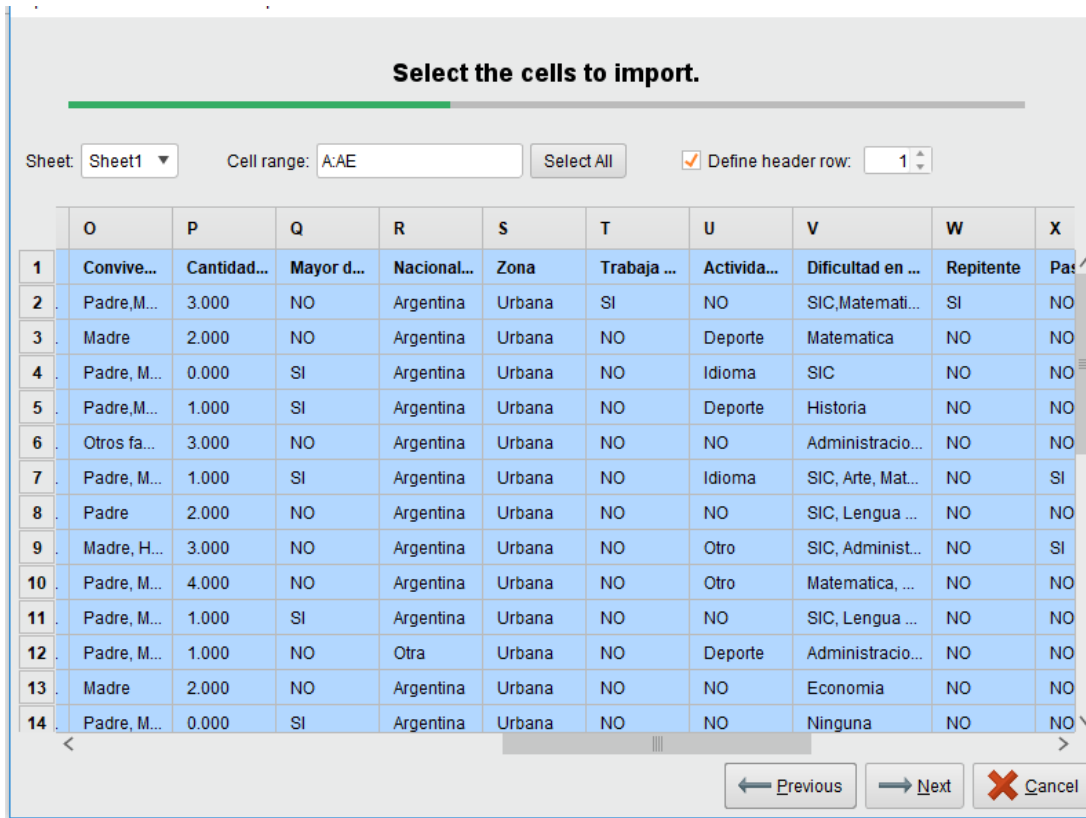


Imagen 17-Importando archivo Excel

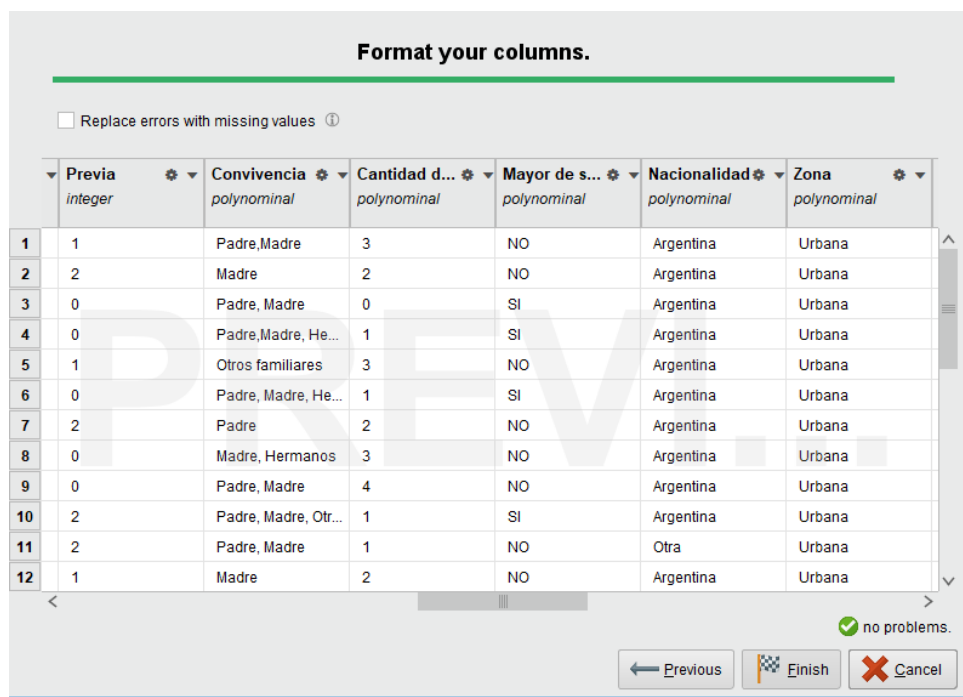


Imagen 18-- Importando archivo Excel (selección de tipo de dato)

Con el software, se pudo analizar con más claridad los datos recolectados. A continuación, se muestran algunos gráficos generados para entender la distribución de los datos

En la sección *Statistics* de RapidMiner, pudimos observar un histograma (Imagen 19), para el campo *Repitentes* (de tipo *polynomial*) en donde tenemos un total de 8 alumnos que son repitentes, mientras que 24 de ellos, no.

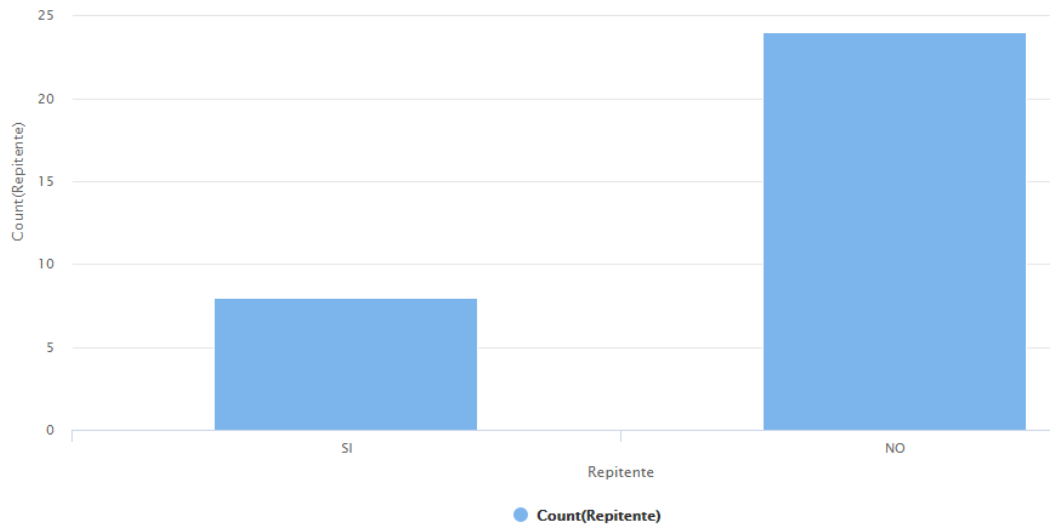


Imagen 19-Cantidad de repitentes y no repitentes

También se puede observar en la Imagen 20 un histograma para el campo *edad* (de tipo *integer*). Este representa la edad de los alumnos, la cual varía entre 15 y 18 años, teniendo mayor cantidad de alumnos con 15 y 16 años, mientras que menores cantidades para aquellos que tienen entre 17 y 18 años.

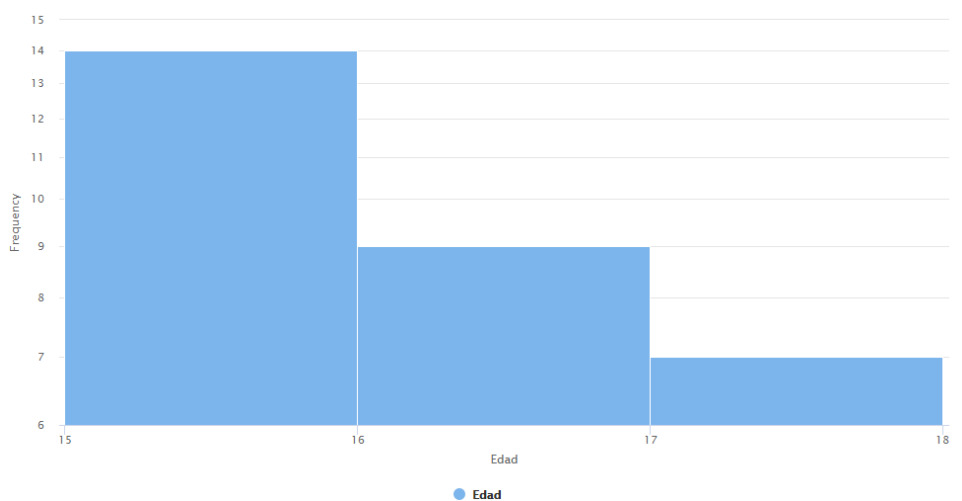


Imagen 20-Histograma para atributo EDAD

La Imagen 21 permite observar que existen 7 alumnos repitentes cuya edad esta entre 17 y 18 años (barra color celeste), 1 alumno repitente de 16 años (barra color azul) y 24 que no son repitentes (barra color verde).

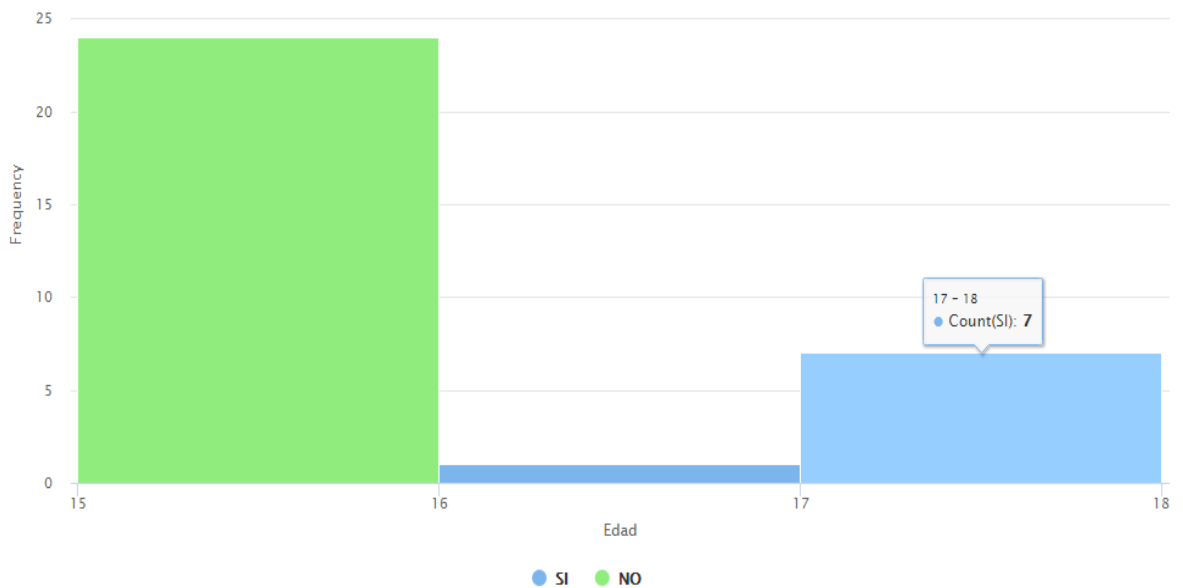


Imagen 21-Edad de alumnos repitentes

Por medio de la Imagen 22, se puede observar que tenemos 2 alumnos que ingresaron con pase a la institución, mientras que los 30 restantes no.

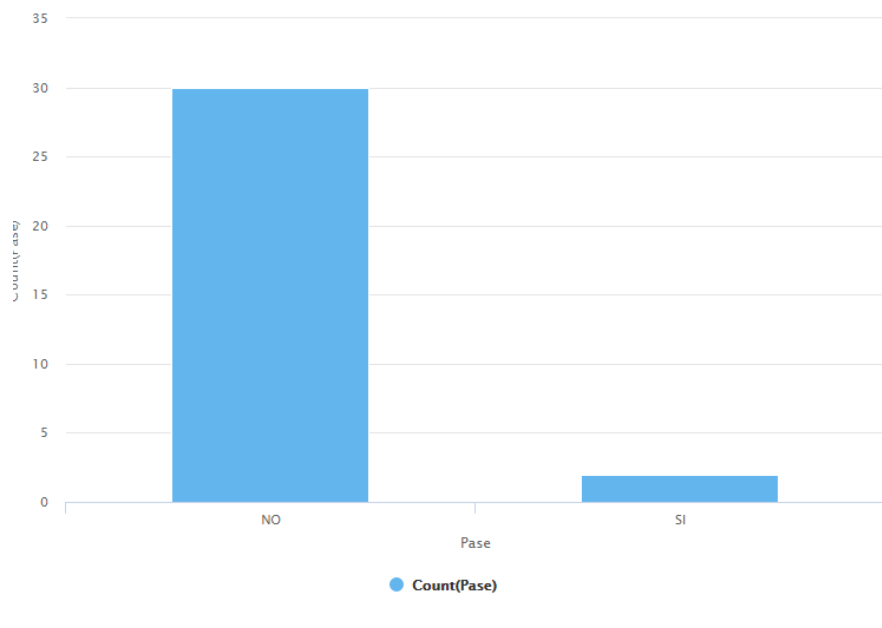


Imagen 22-Histograma para atributo PASE

Con el resto de los atributos (uno por uno) se realizó el mismo procedimiento de análisis mediante gráficos.

▪ **Fase 3: Preparación de los datos:**

Una vez realizado el análisis, se identificaron datos faltantes. Para los 2 alumnos que no completaron la encuesta, existían solo los datos de los registros de calificaciones y asistencias, lo cual representa el 44,83% del total de atributos de cada alumno.

De un total de 29 atributos, 13 corresponden a los datos obtenidos de los registros de notas y asistencias; mientras que 16 atributos corresponden a los datos de las encuestas. Por lo tanto, se tiene un 55,17% de datos faltantes de los dos alumnos que no completaron la encuesta. Por ello, se decidió eliminar las mismas.

Debido a que los alumnos podían seleccionar más de una opción en algunas preguntas de la encuesta, existían columnas que tenían más de un valor en la misma celda como lo muestra la Imagen 23.

| | | |
|--------------------------------|------------------------|---|
| Convivencia | Actividad Extraescolar | Dificultad en materia |
| Padre, Madre | NO | SIC, Matemática, Lengua Extranjera |
| Madre | Deporte | Matemática |
| Padre, Madre | Idioma | SIC |
| Padre, Madre, Hermano | Deporte | Historia |
| Otros familiares | NO | Administración, Lengua Extranjera, ! |
| Padre, Madre, Hermanos | Idioma | SIC, Arte, Matemática |
| Padre | NO | SIC, Lengua Extranjera |
| Madre, Hermanos | Otro | SIC, Administración, Lengua y Litera |
| Padre, Madre | Otro | Matemática, Educación Física |
| Padre, Madre, Otros familiares | NO | SIC, Lengua Extranjera |
| Padre, Madre | Deporte | Administración, Lengua Extranjera |
| Madre | NO | Economía |
| Padre, Madre | NO | Ninguna |
| Padre, Madre, Hermanos | NO | SIC, Administración, Economía |
| Madre | NO | SIC, Matemática, Lengua Extranjera |
| Padre, Madre, Otros familiares | NO | Ninguna |
| Padre, Madre | Idioma | SIC, Educación Física, Arte, Matemática |
| Padre, Madre | Otro | Ninguna |
| Padre, Madre | Idioma, Deporte | Educación Física |
| Padre, Madre | Deporte | Matemática, Lengua Extranjera |
| Madre | NO | Economía, Administración, Lengua y |

Imagen 23-Campos con más de un valor

La columna “*Dificultad en materia*” (la cual hace referencia a la pregunta “¿*Qué materia/s te costaron más en tercer año?*” de la Imagen 9) posee valores múltiples en las celdas como ser: *Matemática, Lengua, Contabilidad*, etc. Sucediendo lo mismo con los valores de la columna *Convivencia* y *Actividad Extraescolar*. (Imagen 23)

Esto hace difícil que el algoritmo pueda trabajar de manera correcta.

Por tal razón, se decidió modificar el valor de estos atributos, por uno nuevo como lo muestra la Tabla 1. Además, se modificó el nombre y valor de los siguientes atributos: *Distancia casa-institución, Nacionalidad, Zona*; con el fin de lograr que éstos sean más comprensibles.

También, se modificó el nombre y el valor de los atributos *Cantidad de hnos* y *Tiempo dedicado a estudios*, ya que pudimos observar en sus respectivos diagramas de dispersión (Imagen 24 e Imagen 25) que algunos valores se encontraban dispersos.

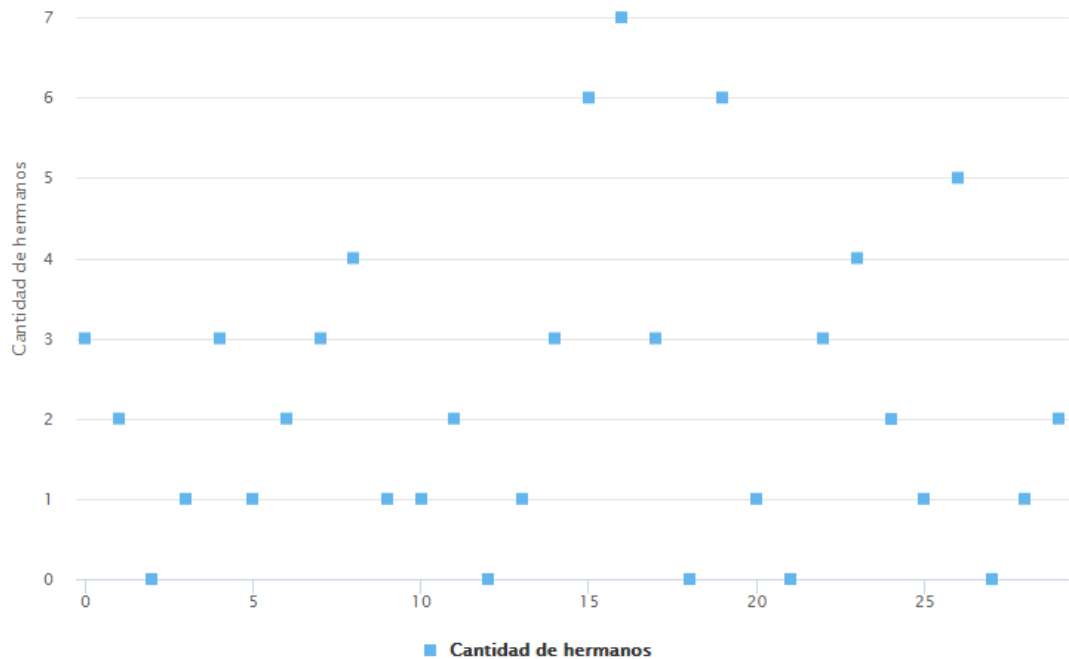


Imagen 24- Valores dispersos para campo CANT. DE HERMANOS

Se observó en la Imagen 24, que, en la primera línea horizontal (empezando desde abajo) 5 alumnos no tienen hermanos; en la segunda línea horizontal 8 alumnos tienen 1 solo hermano, luego en la cuarta línea, 6 alumnos con 3 hermanos y en la octava línea se observa 1 solo alumno con 7 hermanos.

Por otro lado, en la Imagen 25 también se observaron datos dispersos para el atributo *Tiempo dedicado a estudios*, ya que tenemos 16 alumnos que le dedican al estudio entre 1 y 2 horas, mientras que solo tenemos 4 que dedican más de 3 horas, y 3 que dedican más de 1 hora.

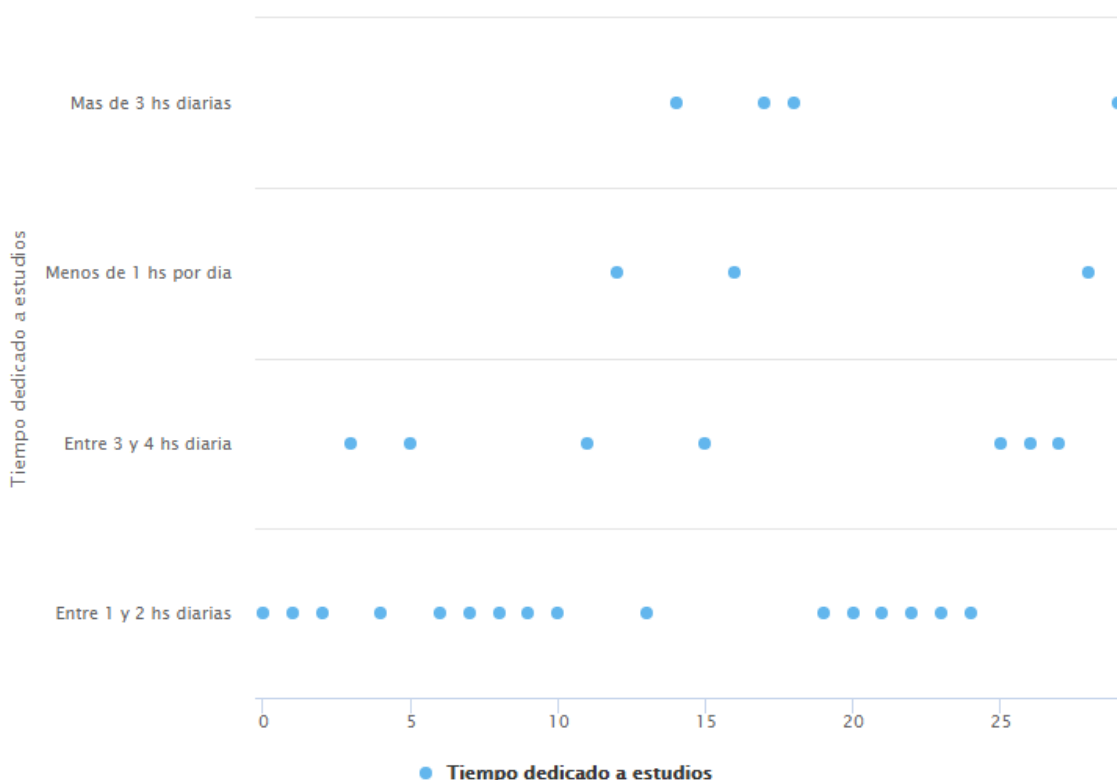


Imagen 25-Valores dispersos para campo TIEMPO DEDICADO A LOS ESTUDIOS

Los datos dispersos pueden provocar que el algoritmo no obtenga buenos resultados. Por ello, y con el fin de obtener un conjunto de datos y un modelo de minería de datos de calidad, se decidió agrupar los valores y transformarlos en datos equivalentes. (Lin, 2002)

La Tabla 1 muestra lo mencionado.

| Atributo original | Atributo nuevo | Valor |
|----------------------------|-------------------------------|---------|
| Convivencia | Vive con ambos padres | SI- NO |
| Dificultad en materia | Sin cambios | SI-NO |
| Cantidad de hnos | Más de 2 hnos | SI- NO |
| Distancia casa-institución | Vive más de 10 cuadras | SI - NO |
| Nacionalidad | Nacionalidad Argentina | SI-NO |
| Zona | Vive zona urbana | SI-NO |
| Actividad Extraescolar | Sin cambios | SI-NO |
| Tiempo dedicado a estudios | Dedica más de 2 hs al estudio | SI-NO |

Tabla 1- Modificación de atributos

Se prosiguió con la elección del atributo clase. En nuestro caso agregamos una columna denominada “Riesgo académico” cuyos valores posibles son SI y NO.

Según el ministerio de educación, los alumnos que obtengan más de dos materias previas repetirán de curso.

Por lo tanto, deducimos que aquellos alumnos, de acuerdo al registro de notas, que poseían una o dos materias previas están en riesgo escolar.

Además según el capítulo III, art 9 del Régimen de inasistencias y reincorporaciones, los alumnos pierden por primera vez su regularidad al efectuarse las 20 insistencias. (DAD- Departamento de Aplicación Docente, 2018). Por lo tanto, aquellos alumnos que tengan más del 50% de insistencias, es decir 10, estarán también en situaciones de riesgo.

En base a lo mencionado y de acuerdo a nuestros registros, se detectaron un total de 20 alumnos en riesgo académico, como se muestra en la Imagen 26

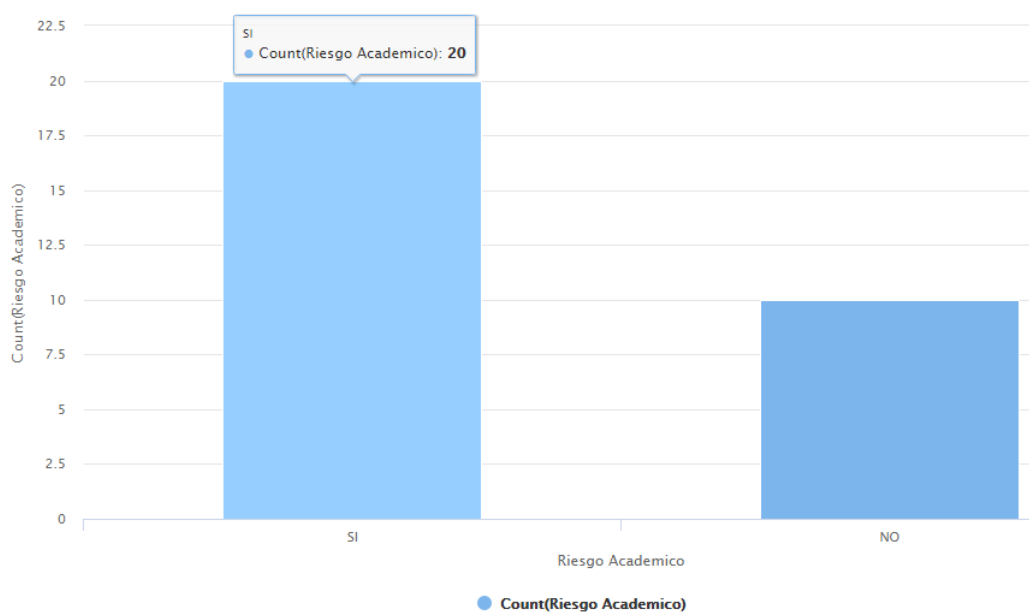


Imagen 26-Histograma para atributo clase (RIESGO ACADÉMICO)

Cabe aclarar, que los datos usados para definir la clase (calificaciones, cantidad de materias previas y cantidad de inasistencias) no intervienen en el modelado.

Además, se excluyeron, de los datos finales, las calificaciones y cantidad de materias previas debido a que se consideró que las notas no aportan información relevante para el descubrimiento de conocimiento oculto, ya que es lógico que los alumnos con materias desaprobadas y con una cantidad elevada de inasistencias, están es riesgo.

Por ello, se procedió a separar las notas e inasistencias de la información recolectada en las encuestas.

Además, se decidió excluir del conjunto de datos, el atributo *Edad*, ya que, de acuerdo a investigaciones previas detalladas en la sección 1.5, éste no forma parte de los posibles factores que influyan en el rendimiento académico.

La Tabla 2 describe los atributos, el tipo y el valor con los que finalmente se creó la vista minable:

| Atributo | Descripción | Tipo | Valor |
|---------------------------------|---|---|--------------------------|
| Vive con ambos padres | Si vive con ambos padres, incluye a hermanos | Polynomial | SI-NO |
| Más de dos hnos | Si tiene más de dos hermanos | Polynomial | SI- NO |
| May de hnos | Si es el mayor de sus hermanos | Polynomial | SI-NO |
| Nacionalidad Arg | Si es de nacionalidad argentina. | Polynomial | SI-NO |
| Vive Zona Urbana | En qué zona vive | Polynomial | SI-NO |
| Trabaja Alumno | Si el alumno trabaja. | Polynomial | SI-NO |
| Actividad extraescolar | Si realiza alguna actividad fuera del horario escolar | Polynomial | SI-NO |
| Dificultad en materia | Si le costaron materias | Polynomial | SI-NO |
| Repitente | Si repitió de curso | Polynomial | SI-NO |
| Pase | Si viene de pase de otro establecimiento | Polynomial | SI-NO |
| Dedica más de dos hs al estudio | Si dedica más de dos horas al estudio | Polynomial | SI-NO |
| Plan/beca | Si es beneficiario de algún plan o beca | Polynomial | SI-NO |
| Vive más de 10 cuadras | Si vive más lejos de 10 cuadras | Polynomial | SI-NO |
| Movilidad | En que medio llega a la institución | Polynomial | Colectivo-Caminando-Otro |
| Tutor | Quien es el tutor | Polynomial | Madre-Padre- otro |
| Trabaja Tutor | Si el tutor trabaja | Polynomial | SI-NO |
| Riesgo académico | Si el alumno se encuentra en riesgo o no. | Polynomial Label (Propiedad que se define para el atributo Clase) | SI-NO |

Tabla 2- Atributos que conforman la vista minable

▪ **Fase 4: Modelado:**

El modelo que se utiliza es de tipo predictivo y la tarea de minería de datos que más se adecua es la de clasificación. Por ello, los algoritmos de clasificación utilizados en esta fase son: Naive Bayes, W-OneR, Decision Tree y WJ48. Cabe aclarar, que se decidió utilizar estos algoritmos debido a que las investigaciones previas detalladas en la sección 1.8, obtuvieron resultados confiables en las predicciones, y, además son frecuentemente utilizados en modelos predictivos de otras áreas.

En RapidMiner, se realizaron una serie de experimentos con los algoritmos mencionados a fin de analizar el comportamiento y verificar qué tan bien pueden predecir. Esto se explica con más nivel de detalle en el capítulo 4.

Como se dispone de 30 registros, se utilizó el método de **validación cruzada** para los algoritmos. Parte del proceso se muestra en la Imagen 27.

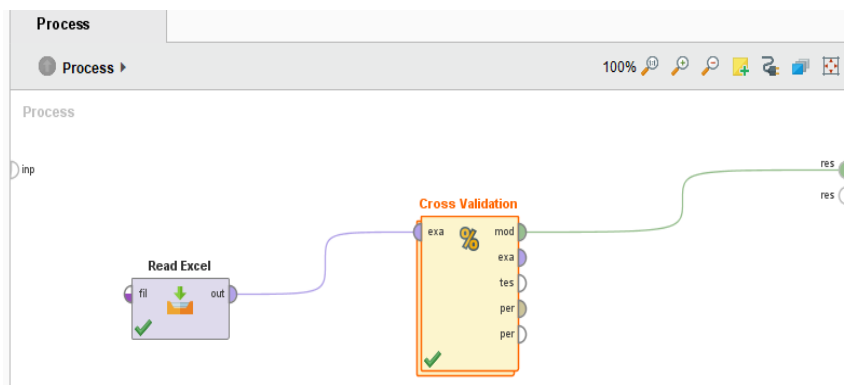


Imagen 27-Proceso de validación cruzada

Las pruebas realizadas con cada algoritmo durante esta etapa se describen con detalle en el capítulo 4.

▪ **Fase 5: Evaluación:**

Una vez realizada la fase de modelado, se prosiguió a evaluar los resultados obtenidos y a elegir aquel algoritmo que tuviera mayor precisión.

En la Tabla 3 se puede observar la precisión y el recall obtenido por cada algoritmo.

| Algoritmo | Precisión | Recall |
|-------------------|-----------|----------------------|
| Árbol de decisión | 66.67% | SI = 85% NO = 30% |
| Naive Bayes | 76,67% | SI = 90% NO= 50% |
| WJ48 | 76,67% | SI = 90% NO= 50% |
| W-OneR | 80,00% | SI=95% NO=50% |

Tabla 3- Precisión y recall por algoritmo

Se valora el recall debido a que es necesario identificar cuantas instancias clasifica el modelo correctamente.

Por otro lado, se puede observar en la Tabla 3 que las precisiones más altas corresponden a los algoritmos WJ48 Y W-OneR. Sin embargo, el algoritmo W-OneR fue considerado ineficiente y descartado debido a que devolvió, en cada

prueba realizada (ver capítulo 4 - Imagen 37, Imagen 40, Imagen 52), una predicción basada en un solo atributo, sin tener en cuenta los demás.

Con el algoritmo WJ48 se obtuvo el mismo resultado que el algoritmo W-OneR; es decir que, en cada prueba realizada (ver capítulo 4 - Imagen 35, Imagen 44, Imagen 49), se obtuvo un árbol con solo un atributo y una sola regla. Por tal razón, también se determinó ineficiente y se descartó.

Finalmente, el algoritmo elegido como *el mejor* es *Naive Bayes*, debido a la mayor precisión en comparación con el algoritmo Decision Tree.

Luego de realizar este análisis, comprender los resultados obtenidos y seleccionar el mejor algoritmo, pudimos avanzar a la siguiente y última fase del proceso.

▪ **Fase 6: Despliegue:**

El modelo obtenido fue incorporado al sistema de alarma para que los usuarios puedan utilizarlo en el momento que se requiera y visualizar el riesgo académico, sin necesidad de tener conocimientos técnicos sobre minería de datos. Esto facilita el entendimiento y la interacción de los actores educativos con el sistema.

Cabe aclarar que el resultado de la predicción del modelo mostrará la “Probabilidad de riesgo por factores externos” posterior a una carga de datos de notas, insistencias y datos personales del alumno por trimestre.

3.3. Funcionamiento interno de la solución propuesta

El funcionamiento interno de la solución se muestra en la Imagen 28

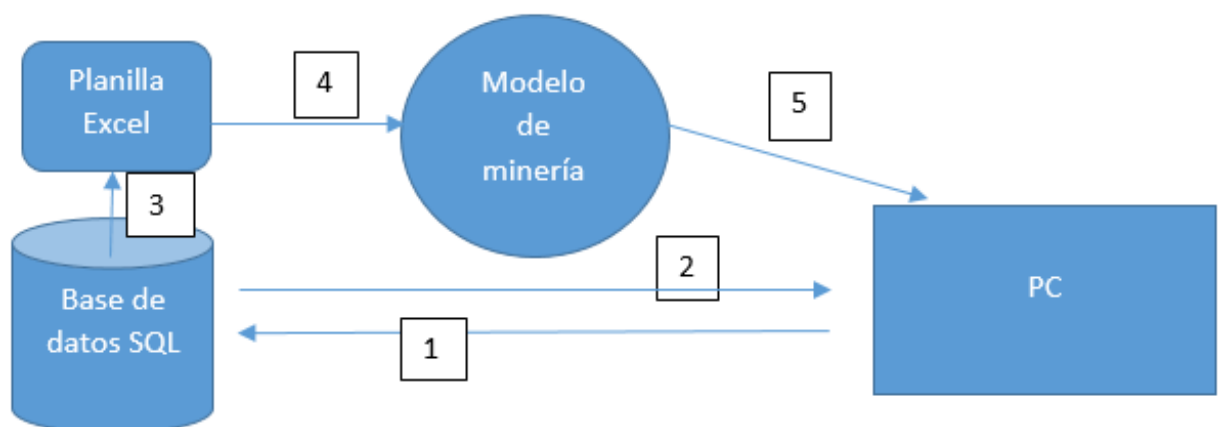


Imagen 28-Funcionamiento interno de la solución

- 1- Al realizar una carga de datos (notas, inasistencias e información personal) desde la computadora, éstos se almacenan en la base de datos SQL Server.
- 2- Las notas e inasistencias almacenadas son utilizadas para determinar el riesgo por notas o inasistencias y será mostrado al usuario.
- 3- Mientras que la información personal de los alumnos, también almacenada en la base de datos, es usada para crear una planilla Excel con extensión .csv de manera automática.
- 4- A esta planilla generada, la usa el modelo de minería de datos para determinar la predicción y el resultado de la predicción es devuelto en un archivo Excel.
- 5- Los resultados finalmente son mostrados al usuario en la computadora.

3.4. Análisis FODA

3.4.1. Fortalezas

- Se cuenta con el asesoramiento de personas con conocimientos en minería de datos.
- Experiencia previa en desarrollo de prototipos de sistemas.
- Herramientas de desarrollo de software y de minería de datos gratuitas.
- Disposición de los actores educativos para llevar a cabo proyecto.

3.4.2. Debilidades

- Acceso a información de un solo curso y división de la institución.
- Tecnología institucional desactualizada.
- Escaso seguimiento de la situación escolar (notas e insistencia) del alumnado.

3.4.3. Oportunidades

- La solución puede extenderse a todo el establecimiento.
- Mejorar la calidad educativa de la institución.
- Disminuir el bajo rendimiento académico.
- Disponibilidad de la información en el momento en que se requiera.
- Rápida recuperación de información.

3.4.4. Amenazas

- Resistencia al cambio por parte de los actores institucionales frente a las nuevas tecnologías.

3.5. Análisis de factibilidad

3.5.1. Factibilidad económica

Las herramientas de desarrollo de software utilizadas para la solución propuesta no incluyen ningún costo. Las mismas se detallan en la Tabla 4.

| Ítem | Precio | Descripción |
|-------------------------|--------|-----------------------------|
| RapidMiner Studio | \$0 | Bajo licencia estudiantil. |
| Visual Studio Community | \$0 | Versión Community gratuita. |
| SQL Server 2012 | \$0 | Versión Express gratuita. |
| Total | \$0 | |

Tabla 4- Herramientas de desarrollo y costos asociados

A pesar de lo mencionado, para el uso del sistema en la institución, es necesario efectuar el pago de las correspondientes licencias. Las mismas se detallan en la Tabla 5.¹⁶

| Software | Precio | Descripción |
|-------------------------|---------------------------|-----------------------------|
| RapidMiner | \$56.000 (\$1.000,00 USD) | Licencia anual. |
| Visual Studio Community | \$0 | Versión Community gratuita. |
| SQL Server | \$0 | Versión Express gratuita. |
| Total | \$5000 anual. | - |

Tabla 5- Licencias Software

Por otro lado, la infraestructura tecnológica que posee la institución es antigua. Lo que produce una consecuencia importante.

La escuela cuenta con 4 computadoras de escritorio cuyas características son:

- Procesador Intel Pentium

¹⁶ Términos y condiciones de Visual Studio Community- 08/2018.

<https://visualstudio.microsoft.com/es/license-terms/mlt553321/?rr=https%3A%2F%2Fwww.google.com.pe%2F>

SQL Server Express - 07/2019- <https://microsoft-sql-server.uptodown.com/windows>

RapidMiner Studio - 07/2019- <https://rapidminer.com/products/studio/>

- Memoria RAM 2GB
- Disco duro de 80GB

Ante la imposibilidad de adquirir nuevos equipos se propone actualizar algunos componentes para mejorarlos.

Se detalla en la Tabla 6 los costos de hardware asociados para una computadora.

- Hardware ¹⁷

| Ítem | Cantidad | Precio individual |
|--|----------|-------------------|
| Memoria RAM Ddr 3 de 8 GB- 1600MHz. | 1 | \$2500,00 |
| Disco SSD: 240 GB SATA 3- 2.5 pulgadas | 1 | \$2300,00 |
| Procesador Intel Core i3- 8100 – 3,5 GHz | 1 | \$5500,00 |
| TOTAL | | \$10.300,00 |

Tabla 6- Hardware necesario y costos asociados

El desarrollo de la solución involucra a los siguientes profesionales:

- RRHH¹⁸

| Ítem | Precio individual | Cantidad de Horas | Total Ítem |
|-------------|-------------------|-------------------|--------------|
| Programador | \$ 200,00 | 200 | \$ 40.000,00 |
| Analista | \$ 300,00 | 40 | \$ 12.000,00 |
| TOTAL | | | \$ 52.000,00 |

Tabla 7- Recursos humanos necesarios y costos asociados

Inversión total (Tabla 8).

| Ítem | Precio |
|----------|-----------------------|
| Hardware | \$ 10.300,00 |
| Software | \$ 56.000,00 (anual). |
| RRHH | \$ 52.000,00 |
| TOTAL | \$123.300,00 |

Tabla 8- Inversión total

La institución no recibe aportes económicos desde el Ministerio de Educación de la provincia para la implementación del sistema software, según manifestaciones de la máxima autoridad de la institución; quien expresa estar interesada en el uso de la

¹⁷ Mercado Libre- 06/2019

¹⁸ Página Oficial COPAIPA- 09/2018

tecnología ofrecida, porque considera que es una herramienta tecnológica que podría mejorar la calidad educativa. Por ello, el sistema software sería financiado con fondos propios que se obtendrían de realización de rifas, quiosco, inscripción por alumnos, inscripción de exámenes, eventos que realiza la institución: bailes escolares (elección reina, aniversario, y otros).

La forma de pago del sistema es la siguiente: se deberá abonar el 50% al inicio del proyecto y el 50% restante al finalizar el mismo.

3.5.2. Factibilidad técnica

Es factible técnicamente dicha solución ya que las herramientas que se utilizaron para la misma se adecúa en cualquier estación de trabajo.

3.5.3. Factibilidad legal

Legalmente, la solución mencionada no infringe ninguna ley. Los datos personales de los alumnos no fueron incluidos en el modelado y se resguardó su identidad en todo momento. (Convenio de derechos del niño)¹⁹

El personal directivo, al ser la máxima autoridad de la institución es quien autoriza la instalación del sistema, cuando considera de utilidad para toda la institución acorde a las necesidades, siempre a favor del educando y de la calidad educativa. (Ley de Educación N° 26.206)²⁰

Por lo tanto, la solución (sección 3.2) se considera factible legalmente.

3.6. Análisis de riesgos

Posterior al análisis de posibles riesgos que pueden surgir en el proyecto, se identificaron riesgos no potenciales pero relevantes. Los mismos se detallan en la Tabla 9.

¹⁹ Convenio de los derechos del niño: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/0-4999/249/norma.htm>

²⁰ Ley de Educación N° 26.206. <https://www.argentina.gob.ar/sites/default/files/ley-de-educ-nac-58ac89392ea4c.pdf>

| Riesgo | Probabilidad | Porcentaje de Importancia | Acción a realizar |
|--|--------------|---------------------------|---|
| Alumnos brinden datos falsos. Por lo tanto podría existir una predicción de riesgo escolar equivocada. | 50% | 100% | Comprobar la veracidad de la información brindada por medio de la colaboración de padres y/o tutores. |
| Actualización de las herramientas que conlleve a la incompatibilidad. | 30% | 10% | Buscar e instalar versiones compatibles. |

Tabla 9- Análisis de riesgo

Capítulo 4: Pruebas y resultados

Para obtener un modelo de minería de datos de calidad, se han realizado múltiples pruebas que se detallarán a continuación.

El conjunto de datos usado al inicio estuvo formado por los atributos de la Tabla 2.

En RapidMiner, el proceso de modelado *Cross Validation* se muestra en la Imagen 29

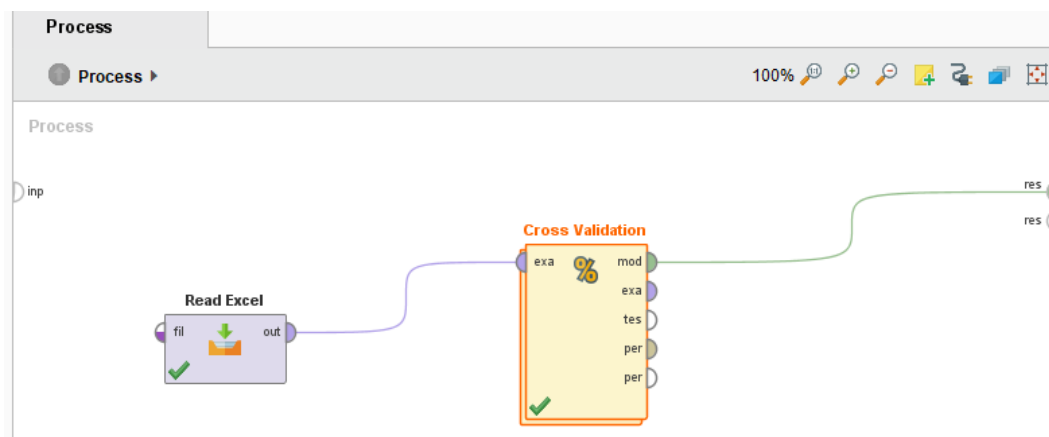


Imagen 29-Proceso de modelado Cross Validation

El operador Cross validation, es un operador anidado. Tiene dos subprocesos: un subproceso de entrenamiento y un subproceso de prueba, como se observa en la Imagen 30. El subproceso Training se usa para entrenar un modelo. El modelo entrenado se aplica luego en el subproceso de prueba (Testing). El rendimiento del modelo se mide durante la fase de prueba.

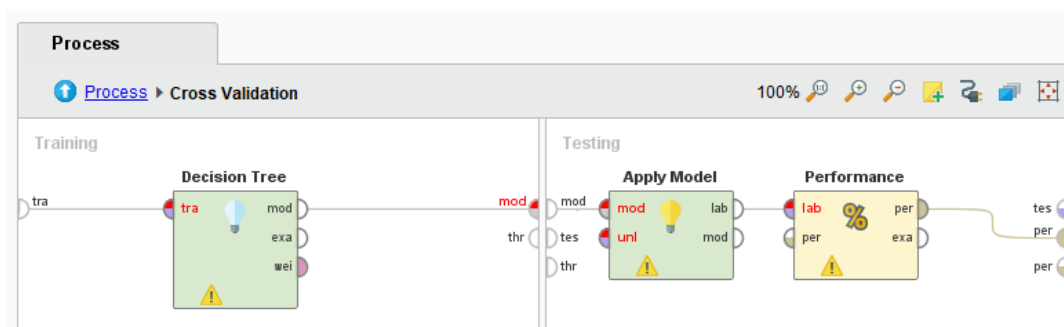


Imagen 30-Subproceso cross validation

En primer lugar, se probaron los algoritmos con todos los atributos disponibles, luego, se seleccionaron los atributos más relevantes, de acuerdo a los operadores de RapidMiner. En todos los casos, se ejecutó el proceso de entrenamiento para los cuatro algoritmos descritos anteriormente: Naive Bayes, Decision Tree, WJ48 y W-OneR. A continuación, se muestran las matrices de confusión y el modelo obtenido de cada entrenamiento.

4.1. Prueba N° 1.

Decision Tree.

Árbol de decisión obtenido (Imagen 31).

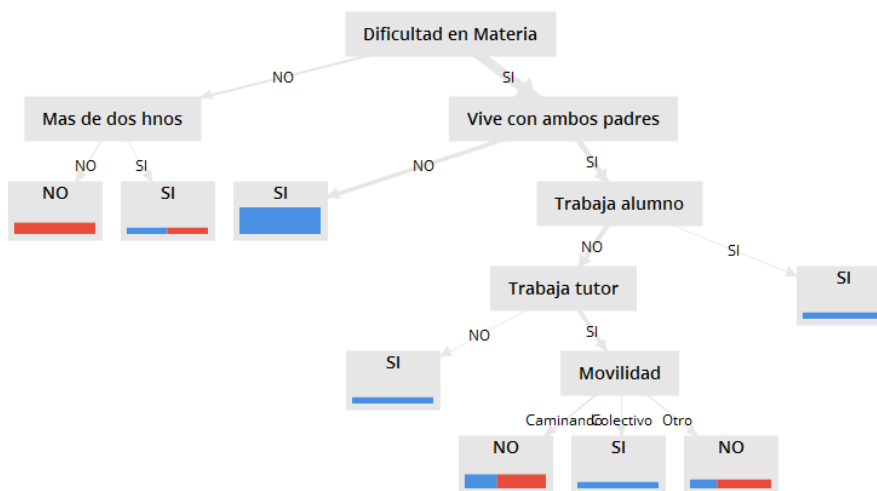


Imagen 31-Árbol de decisión obtenido

La matriz de confusión se muestra en la Imagen 32

accuracy: 63.33% +/- 24.60% (micro average: 63.33%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 16 | 7 | 69.57% |
| pred. NO | 4 | 3 | 42.86% |
| class recall | 80.00% | 30.00% | |

Imagen 32-Matriz de confusión del árbol de decisión

Naive Bayes.

Al probar con el algoritmo de Naive Bayes, se obtuvo la matriz de confusión de la Imagen 33.

accuracy: 66.67% +/- 27.22% (micro average: 66.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 16 | 6 | 72.73% |
| pred. NO | 4 | 4 | 50.00% |
| class recall | 80.00% | 40.00% | |

Imagen 33-Matriz de confusión del algoritmo Naive Bayes

Wj48.

Para WJ48, se obtuvo la matriz de confusión de la Imagen 34.

accuracy: 76.67% +/- 27.44% (micro average: 76.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 18 | 5 | 78.26% |
| pred. NO | 2 | 5 | 71.43% |
| class recall | 90.00% | 50.00% | |

Imagen 34-Matriz de confusión del algoritmo WJ48

W-J48

J48 pruned tree

Dificultad en Materia = SI: SI (24.0/5.0)

Dificultad en Materia = NO: NO (6.0/1.0)

Number of Leaves : 2

Size of the tree : 3

Imagen 35- Descripción de algoritmo WJ48

W-OneR.

Para W-OneR, se obtuvo la matriz de confusión de la Imagen 36.

accuracy: 80.00% +/- 23.31% (micro average: 80.00%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 19 | 5 | 79.17% |
| pred. NO | 1 | 5 | 83.33% |
| class recall | 95.00% | 50.00% | |

Imagen 36-Matriz de confusión del algoritmo W-OneR

W-OneR

```
Dificultad en Materia:  
SI      -> SI  
NO      -> NO  
(24/30 instances correct)
```

Imagen 37-Descripción de algoritmo W-OneR

4.2. Prueba N° 2: Operador Chi-squared.

Luego, se realizaron pruebas con otros operadores de RapidMiner como ser *Chi Squared* a fin de evaluar por completo cada uno de los algoritmos en busca de precisiones más altas.

El operador *Chi-Squared* calcula la relevancia de los atributos para cada atributo de la entrada basándose en el estadístico chi cuadrado.

Proceso en Imagen 38

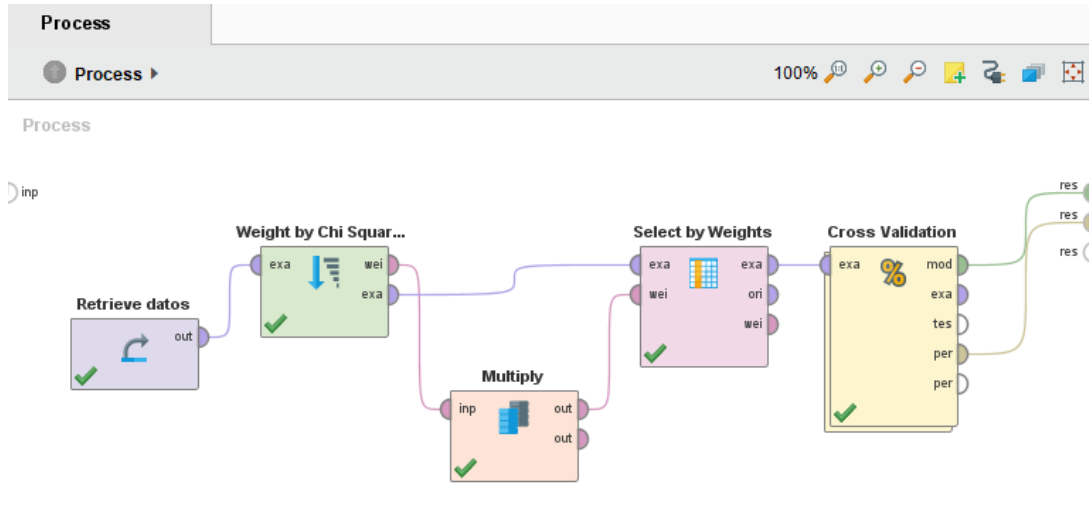


Imagen 38--Proceso con CHI-SQUARED

W-OneR - Matriz de confusión en Imagen 39

accuracy: 80.00% +/- 23.31% (micro average: 80.00%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 19 | 5 | 79.17% |
| pred. NO | 1 | 5 | 83.33% |
| class recall | 95.00% | 50.00% | |

Imagen 39- W-OneR con el operador CHI-SQUARED

W-OneR

Dificultad en Materia:
 SI -> SI
 NO -> NO
 (24/30 instances correct)

Imagen 40-Descripción de algoritmo W-OneR

Decision Tree- Matriz de confusión en Imagen 41

accuracy: 66.67% +/- 22.22% (micro average: 66.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 17 | 7 | 70.83% |
| pred. NO | 3 | 3 | 50.00% |
| class recall | 85.00% | 30.00% | |

Imagen 41-Matriz de confusión de Decision Tree con el operador CHI-SQUARED

Árbol de decisión obtenido. Imagen 42

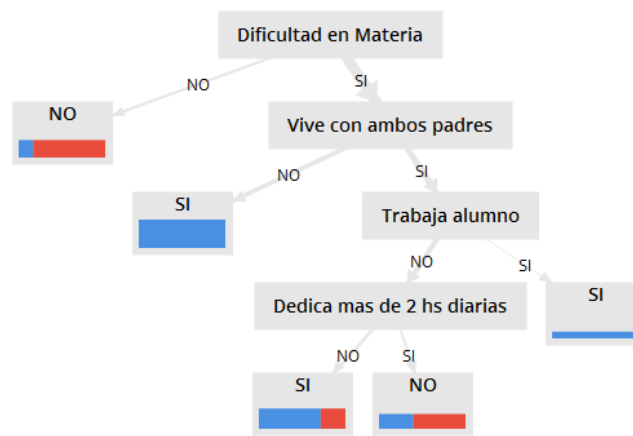


Imagen 42-Árbol de decisión

WJ48- Matriz de confusión en Imagen 43

accuracy: 76.67% +/- 27.44% (micro average: 76.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 18 | 5 | 78.26% |
| pred. NO | 2 | 5 | 71.43% |
| class recall | 90.00% | 50.00% | |

Imagen 43-Matriz de confusión de árbol de decisión con el operador CHI-SQUARED

W-J48

```
J48 pruned tree
-----
Dificultad en Materia = SI: SI (24.0/5.0)
Dificultad en Materia = NO: NO (6.0/1.0)

Number of Leaves :    2

Size of the tree :    3
```

Imagen 44-Descripción de algoritmo WJ48

Naive bayes- Matriz de confusión en Imagen 45

accuracy: 76.67% +/- 22.50% (micro average: 76.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 18 | 5 | 78.26% |
| pred. NO | 2 | 5 | 71.43% |
| class recall | 90.00% | 50.00% | |

Imagen 45- Matriz de confusión de Naive Bayes

4.3. Prueba N° 3: Operador Remove Useless Attributes.

El operador *Remove Useless Attributes* remueve aquellos atributos que no son de utilidad para el algoritmo. A continuación, se muestra la prueba realizada con el operador mencionado.

Proceso en Imagen 46.

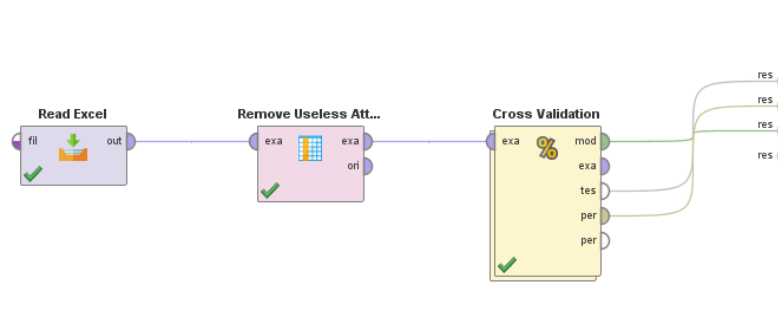


Imagen 46-Proceso con operador REMOVE USSLES

Árbol de Decisión - Imagen 47

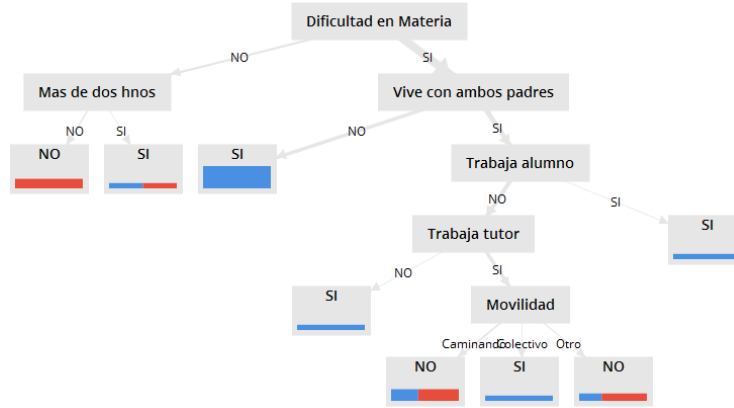


Imagen 47-Árbol de decisión obtenido con REMOVE USSLES

Matriz de confusión en Imagen 48

accuracy: 63.33% +/- 24.60% (micro average: 63.33%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 16 | 7 | 69.57% |
| pred. NO | 4 | 3 | 42.86% |
| class recall | 80.00% | 30.00% | |

Imagen 48-Matriz de confusión de árbol de decisión con operador REMOVE USSLES

Wj48- Matriz de confusión en Imagen 49

accuracy: 76.67% +/- 27.44% (micro average: 76.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 18 | 5 | 78.26% |
| pred. NO | 2 | 5 | 71.43% |
| class recall | 90.00% | 50.00% | |

Imagen 49-Matriz de confusión de WJ48 con operador REMOVE USSLES

W-J48

```
J48 pruned tree
-----

Dificultad en Materia = SI: SI (24.0/5.0)
Dificultad en Materia = NO: NO (6.0/1.0)

Number of Leaves :    2

Size of the tree :    3
```

Imagen 50-Descripción de algoritmo WJ48

W-OneR- Matriz de confusión en Imagen 51

accuracy: 80.00% +/- 23.31% (micro average: 80.00%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 19 | 5 | 79.17% |
| pred. NO | 1 | 5 | 83.33% |
| class recall | 95.00% | 50.00% | |

Imagen 51-Matriz de confusión de W-OneR con operador REMOVE USSLES

W-OneR

```
Dificultad en Materia:
      SI      -> SI
      NO      -> NO
(24/30 instances correct)
```

Imagen 52-Descripción de algoritmo W-OneR

Naive Bayes- Matriz de confusión en Imagen 53

accuracy: 66.67% +/- 27.22% (micro average: 66.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 16 | 6 | 72.73% |
| pred. NO | 4 | 4 | 50.00% |
| class recall | 80.00% | 40.00% | |

Imagen 53- Matriz de confusión de Naive Bayes con operador REMOVE USSLES

A continuación, se muestra en la Tabla 10 un resumen de las precisiones obtenidas por algoritmo en cada prueba.

| Algoritmo | Prueba 1. | Prueba 2: Operador Chi-squared. | Prueba 3: Operador Remove Useless Attributes. |
|---------------|-----------|------------------------------------|--|
| W-OneR | 80,00 % | 80,00 % | 80,00 % |
| WJ48 | 76,67 % | 76,67 % | 76,67 % |
| Decision Tree | 63,33 % | 66,67 % | 63,33 % |
| Naive Bayes | 66,67 % | 76,67 % | 66,67 % |

Tabla 10-Resumen de precisiones obtenidas en cada prueba

Finalmente, dadas las pruebas realizadas y las precisiones obtenidas por cada algoritmo, se determinó que el mejor algoritmo de clasificación es el de Naive Bayes de la prueba 2 (con operador Chi squared), cuya precisión es del 76,67%. Es necesario recordar que los algoritmos WJ48 y W-OneR fueron descartados debido a la desconfianza que presentaban al obtener una predicción basada en un solo atributo.

Como vemos en la Imagen 54, el algoritmo seleccionado predijo a 18 alumnos en situación de riesgo que en realidad se encontraban en situación de riesgo, mientras que predijo 2 alumnos que no estaban en riesgo, pero en realidad si lo estaban. Por otro lado, predijo un total de 5 alumnos en riesgo que en realidad no lo estaban, mientras que predijo a 5 de ellos, que no estaban en riesgo y que realmente no lo estaban.

accuracy: 76.67% +/- 22.50% (micro average: 76.67%)

| | true SI | true NO | class precision |
|--------------|---------|---------|-----------------|
| pred. SI | 18 | 5 | 78.26% |
| pred. NO | 2 | 5 | 71.43% |
| class recall | 90.00% | 50.00% | |

Imagen 54-Matriz de confusión con precisión más alta- Naive Bayes

El recall (90%) obtenido es un buen numero. Esto significa que, de los 20 alumnos en riesgo que debía marcar con SI, clasificó bien a 18 de ellos, es decir, a casi todos. Por otro lado, predijo a 5 alumnos como SI, pero en realidad NO estaban en riesgo, lo cual no sería un error tan grave. Esto quiere decir que, si la escuela hiciera un seguimiento al alumnado en situación de riesgo escolar (según la predicción del modelo), es preferente realizar un seguimiento, aunque el modelo se haya equivocado, en vez de no considerarlos.

4.4. Incorporación del modelo obtenido al sistema de alarma.

En RapidMiner se creó un proceso de clasificación para nuevas instancias usando el modelo entrenado y seleccionado como se muestra en la Imagen 55. Este proceso, recibe un archivo generado por el sistema de alarma cuya extensión es .csv que contendrá nuevos datos (datos de los alumnos que se quiere evaluar) y aplicará el modelo a ese conjunto de datos con el operador *apply model* (Imagen 55).

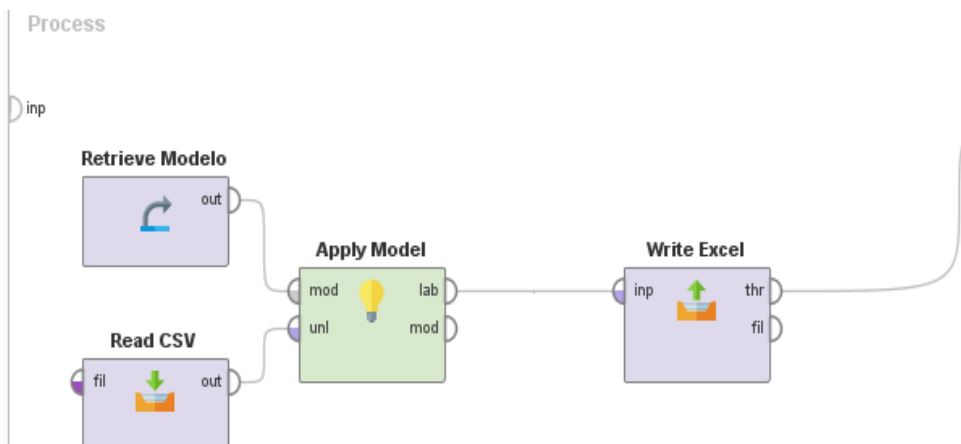


Imagen 55-Proceso que utiliza el modelo generado y un nuevo conjunto de datos para predecir nuevas instancias

Este proceso, es guardado en un repositorio local de RapidMiner para luego ser llamado por el sistema de alarma por medio de líneas de código C# con el fin de clasificar las nuevas instancias, es decir, indicar la probabilidad de riesgo académico. Como resultado de este proceso se genera un archivo Excel que contiene las instancias ya clasificadas y es usado por el sistema para mostrar los resultados en la pantalla de la computadora (Imagen 56).

| | Nombre del Alumno | Notas | Inasistencias | Probabilidad de Riesgo por Factores Externos. | Información. |
|---|-------------------|-------|---------------|---|--------------|
| ▶ | Ruben | | | | |
| | Gaston | | | | |

Referencia.
Alumnos con riesgo ████████

Imagen 56-Resultados

Lo mencionado anteriormente, completa los pasos 4 y 5 (Imagen 28) del proceso presentado como solución propuesta en la sección 3.3.

Conclusión

Durante el desarrollo del trabajo, se logró identificar la importancia de los problemas educativos referentes al bajo rendimiento escolar y a la repitencia escolar; ya que éstos alteran la calidad educativa de la institución y además la trayectoria escolar de los estudiantes.

Mediante la aplicación de la minería de datos a un conjunto de datos de 30 alumnos, se logró anticipar situaciones de riesgo escolar. El porcentaje obtenido del modelo final puede mejorar a medida que se incorporen más datos, es decir, datos de los demás cursos. Por la forma en la que fue desarrollado, con la incorporación de más datos, solo se deberá reentrenar el modelo. Esto quiere decir que no es necesario realizar nuevamente todo el proceso de minería de datos en RapidMiner, sino, incorporar al sistema comandos que permiten reentrenarlo automáticamente cada determinado tiempo.

Con el sistema de alarma desarrollado, es posible realizar un seguimiento de situaciones escolares a fines de anticiparse a los hechos mencionados anteriormente (repitencia y bajo rendimiento).

Con este sistema, los preceptores podrán informar a los padres o tutores de los alumnos, la situación académica en relación al bajo rendimiento de los mismos en el momento en que se lo requiera.

Por otro lado, los docentes podrán obtener un diagnóstico en cada trimestre, con el fin de planificar estrategias superadoras en el desarrollo del proceso enseñanza-aprendizaje.

El sistema informático, es, por ahora, un prototipo que dispone de una alarma de riesgo escolar según calificaciones, inasistencias, y otros factores (usados por el modelo). Sin embargo, en un futuro se pueden agregar otras opciones tales como: acceso al sistema para registrar horarios de docentes, cursos a cargo, registro de programas, informes trimestrales o planillas de alumnos aprobados/desaprobados; o el registro de alumnos que retiraron los títulos al finalizar el secundario, horarios de turnos de exámenes; o registrar a los alumnos que rindieron materias previas o equivalencias. De esta manera, también incrementaría la cantidad de datos (atributos) por cada alumno, es decir, podrían tener datos acerca de profesores que tienen, notas y asistencias de otros trimestres, cantidad de unidades por materia, entre otros.

Algunas de estas opciones, también en un futuro, podrían usarse para otro modelado.

También pueden incluirse a todos los cursos y a todas las divisiones del establecimiento, incluyendo el turno mañana. Con lo mencionado, los actores educativos podrán realizar un seguimiento de los alumnos en general y además podrán utilizar la información con fines estadísticos, como por ejemplo para la actualización/renovación del PEI (Proyecto Educativo Institucional).

En conclusión, este sistema de alarma posee un amplio alcance en cuanto a almacenamiento de información y exposición de la misma a favor de la institución.

Desde lo personal, desarrollar este trabajo me llevo a conocer diferentes campos de la investigación, como ser las problemáticas escolares, y también a profundizar los conocimientos sobre minería de datos para poder resolver diferentes situaciones requeridas por el sistema creado, por ejemplo, la integración del sistema con el modelo de minería obtenido. Durante el desarrollo del trabajo se logró poner en práctica los conocimientos adquiridos en el cursado de la carrera, logrando de esta manera brindar una solución a una problemática actual del ámbito educativo, resultando esto una experiencia gratificante.

Bibliografía

- Aguilar, J. (2013). *Big Data: Analisis de grandes volumnes de datos en organizaciones*. Mexico: ALFAOMEGA.
- Amaya Torrado, K., Avendaño Barrientos, E., & Heredia Vizcaíno, D. (2012). *Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos*. Colombia: Universidad Francisco de Paula y Universidad Somón Bolívar.
- Azevedo, A., & Santos, M. (2008). KDD, SEMMA AND CRISP.DM: A PARALLEL OVERVIEW. *IADIS European Conference Data Mining*, (págs. 182-185). Portugal.
- Carrión Perez, E. (20 de 1 de 2002). *Validación de características al ingreso como predictores del rendimiento academico en la carrera de medicina*. Obtenido de Revista Cubana de Educación Media Superior:
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412002000100001&lng=es&nrm=iso
- Castejón, C., & Pérez, S. (1998). Un modelo causal-explicativo sobre la influencia de las variables psicosociales en el rendimiento académico. *Revista Bordón. Sociedad Española de Pedagogía*, 170-184.
- DAD- Departamento de Aplicación Docente. (15 de 09 de 2018). *Régimen de Inasistencias y Reincorporaciones*. Obtenido de <http://dad.uncuyo.edu.ar/regimen.de-asistencia>
- Daedalus. (13 de 10 de 2018). *Ecolink*. Obtenido de <https://m.econlink.com.ar/gestion-conocimiento/gestion-del-conocimiento>
- Fayyad, U. M. (1996). *Data Mining and Knowledge Discovery: Making Sense out of Data*. USA: IEEE Intelligent System.
- Galindo, J. A., & Garcia, H. (s.f.). *Minería de datos en la Educación*. Madrid-España.
- García, S. V. (2013). *Gestión de la prevención de riesgos laborales en pequeños negocios*. Málaga: IC Editorial.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and Techniques*. USA: Morgan Kaufmann Publisher.
- Haro Rivera, S., Zuñiga Lema, L., Freire Meneses, A., Vera Rojas, L., & Escudero Villa, A. (08 de 06 de 2018). *MÉTODOS DE CLASIFICACIÓN EN MINERÍA*. Obtenido de <http://ceaa.esPOCH.edu.ec:8080/revista.perfiles/Articulos/Perfiles20Art13.pdf>
- Hernandez Orallo, J. (2004). *Introducción a la minería de datos*. Madrid: Prentice Hall.
- Hurtado, C. L. (14 de 09 de 2007). *Evaluación de modelos (II)*. Obtenido de U-Cursos:
https://www.u-cursos.cl/ingenieria/2007/2/CC72J/1/material_docente/
- Konchady, M. (2006). *Text Mining Application Programming*. Boston: Charles River Media.

- La Red Martínez, D., Karanik, M., Giovanni, M., & Scappini, R. (2016). *Modelo predictivo de rendimiento académico utilizando la minería de datos en la UTN-FRRe*. Resistencia, Argentina.
- Lin, T. Y. (2002). *Attribute transformation for data mining I: Theoretical explorations*. California, Berkeley: Guest Editorial.
- López Takeyas, B. (10 de 7 de 2019). *Algoritmo C4.5*. Obtenido de [http://www.itnuevolaredo.edu.mx/takeyas/Apuntes/Inteligencia%20Artificial/Apuntes/tareas_alumnos/C4.5/C4.5\(2005-II-B\).pdf](http://www.itnuevolaredo.edu.mx/takeyas/Apuntes/Inteligencia%20Artificial/Apuntes/tareas_alumnos/C4.5/C4.5(2005-II-B).pdf)
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. Israel: Springer.
- Martinez, & Otero, V. (1996). *Factores determinantes del rendimiento académico en la enseñanza media*. Madrid: Publicaciones del colegio de psicólogos de Madrid.
- Mierswa, I., Wurst, M., Klinberg, R., Scholz, M., & Euler, T. (2006). YALE: Rapid Prototyping for Complex Data Mining Tasks, in Proceedings of the 12th ACM SIGKDD. *International Conference on Knowledge Discovery and Data Mining*.
- Ministerio de Educación. (20 de 09 de 2018). *Ministerio de Educación*. Obtenido de <https://www.argentina.gob.ar/educacion>
- Ministerio de Educación. (21 de 09 de 2018). *Prueba Aprender 2017*. Obtenido de <https://www.argentina.gob.ar/educacion/aprender2017>
- Muelle, L. (15 de 7 de 2016). *Apuntes*. Obtenido de Factores de riesgo en el bajo desempeño académico y desigualdad social en el Perú según PISA 2012: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S0252-18652016000200001
- Núñez, J. C. (1998). Estrategias de aprendizaje, autoconcepto y rendimiento académico. *Revista Psicothema*.
- Orea, S., Vargas, S., & Alonso Garcia, M. (2009). *Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos*. Izuca de Matamoros, Puebla, México.
- Pellegrina García, S., & Casanova, P. (2002). Parenting styles and adolescents academic performance. *Infancia y Aprendizaje*, 127-168.
- Pérez Luño, A., Jerónimo, R., & Sánchez Vasquez, J. (2000). *Análisis exploratorio de las variables que condicionan el rendimiento académico*. España: Universidad Pablo de Olavide: Sevilla.
- Pérez Porto, J., & Gardey, A. (08 de 11 de 2018). *Definición de repitencia escolar*. Obtenido de <https://definicion.de/repitencia/>
- Pérez Porto, J., & Gardey, A. (10 de 10 de 2018). *Definición de Alarma*. Obtenido de <https://definicion.de/alarma/>

- Pérez Porto, J., & Gardey, A. (11 de 10 de 2018). *Definición de alerta*. Obtenido de <https://definicion.de/alerta/>
- Pérez Porto, J., & Gardey, A. (2 de 11 de 2018). *Definición de rendimiento académico*. Obtenido de <https://definicion.de/rendimiento-academico/>
- Peter, N. (2006). *Introduccion a la computacion*. Mc Graw Hill.
- Quiles, J. G. (2017). *Minería de Datos: algoritmos y modelos*. España: Editorial UOC.
- Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. San Mateo, CA.: Morgan Kaufmann.
- Salanova, M., Cifre, E., Grau, R., & Martinez, I. (2005). Antecedentes de la autoeficacia en profesores y estudiantes universitarios: un modelo causal. *Revista de Psicología del trabajo y de las Organizaciones.*, 159-176.
- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*.
- Tan, P., Steinbach, V., & Kumar, V. (2005). *Introducción a Data Mining*. Addison- Wesley.
- Vega, J. C. (21 de 05 de 2015). *Blog de Traducción e Interpretación Institucional*. Obtenido de Rendimiento académico: <https://blogs.upc.edu.pe/blog-de-traducion-e-interpretacion-profesional/noticias/que-es-el-riesgo-academico>
- Vélez Van, M., & Roa, N. (2005). Factors associated with academic performance in medical students. *PSIC- Educación Media*.
- Witten, I., Frank, E., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed*. USA: Morgan Kaufmann.
- Yourdon, E. (1993). *Análisis Estructurado Moderno*. Mexico: Prentice Hall Hispanoamericano.

Anexo I. Sistema de alarma

El desarrollo del prototipo de un sistema de alarma de detección del riesgo escolar permite identificar a aquellos alumnos que se encuentren en riesgo por calificaciones, inasistencias u otros factores.

En la Imagen 57 se muestra la pantalla principal del sistema.



Imagen 57- Página Principal

Como en todo sistema informático, el control de acceso al mismo es de suma importancia, ya que la información que se maneja es de carácter sensible.

Por esta razón, dicho sistema de alarma requiere el ingreso de un nombre de usuario y una contraseña previo a acceder a la información como se muestra en la Imagen 58.



Volver

Iniciar Sesión

Cuenta Institucional [¿Que es esto?](#)



Usuario

Contraseña

Iniciar Sesión

Imagen 58- Inicio de Sesión

El inicio de sesión garantiza que usuarios no pertenecientes a la institución o personal no autorizado acceda al mismo.

Los preceptores, deben cargar datos pertenecientes a los alumnos como ser: notas, asistencias y datos personales: edad, convivencia, constitución familiar, residencia, trabajo, actividad extraescolar, dificultad en las materias, condición laboral de los padres, tutor, horas semanales dedicadas al estudio, distancia desde el hogar hasta el establecimiento, movilidad y si es beneficiario de alguna beca o plan. Es decir, que se deben cargar los datos obtenidos a partir de las encuestas y los que se usaron para el modelado durante el desarrollo del presente trabajo. Esta carga de datos se debe realizar una vez por trimestre para visualizar el riesgo escolar y así posibilitar un seguimiento al alumnado.

Para ello, se cuenta con un menú principal, el cual aparece, posterior al inicio de sesión, como se muestra en la Imagen 59.

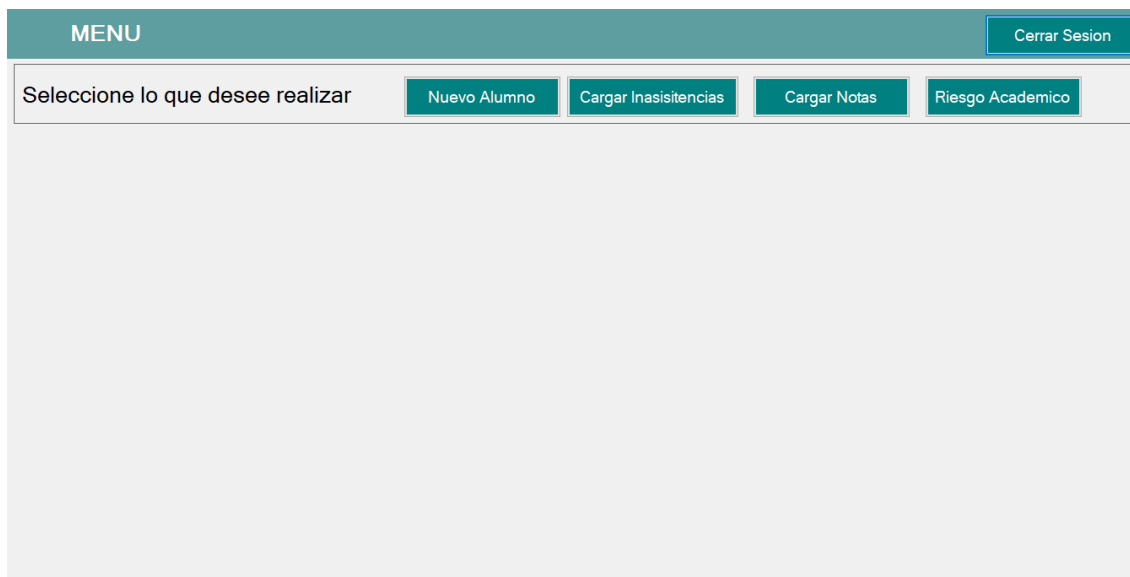


Imagen 59- Menú principal

Si se presiona el botón *Nuevo Alumno* de la Imagen 59, se accede a la sección de carga de datos personales del alumnado (Imagen 60).

The image shows a web application form titled "CARGAR NUEVO ALUMNO". At the top, there is a teal header bar with the title "CARGAR NUEVO ALUMNO" on the left, an "Anterior" button, and a "Cerrar Sesion" button. Below the header, the form is titled "Datos del alumno" and contains several input fields: "Apellidos y Nombres", "DNI", "Edad", "Fecha de Nacimiento", "Direccion", "Telefono", "Curso", and "Division". Each field is represented by a white rectangular input box. At the bottom right of the form, there is a teal button labeled "Siguiete".

Imagen 60- Carga de datos (parte 1)

Si se presiona el botón *Siguiete*, se continúa la carga, como lo muestra la Imagen 61.

| CARGAR NUEVO ALUMNO | | | | Anterior | Cerrar Sesión |
|--|--------------------------|--------------------------|---------------------------------------|--|--------------------------|
| Datos personales | | | | | |
| Vive con ambos padres | <input type="radio"/> SI | <input type="radio"/> NO | Dedica mas de 2 hs diarias al estudio | <input type="radio"/> SI | <input type="radio"/> NO |
| Tiene mas de dos hnos | <input type="radio"/> SI | <input type="radio"/> NO | Posee Plan/Beca | <input type="radio"/> SI | <input type="radio"/> NO |
| Mayor de sus hnos | <input type="radio"/> SI | <input type="radio"/> NO | Vive a mas de 10 cuadras | <input type="radio"/> SI | <input type="radio"/> NO |
| Nacionalidad Argentina | <input type="radio"/> SI | <input type="radio"/> NO | Movilidad | <input type="radio"/> Caminando <input type="radio"/> Colectivo <input type="radio"/> Otro | |
| Vive en Zona Urbana | <input type="radio"/> SI | <input type="radio"/> NO | Quien es el tutor | <input type="radio"/> Padre <input type="radio"/> Madre <input type="radio"/> Otro | |
| El alumno trabaja | <input type="radio"/> SI | <input type="radio"/> NO | El tutor trabaja | <input type="radio"/> SI | <input type="radio"/> NO |
| Actividad Extraescolar | <input type="radio"/> SI | <input type="radio"/> NO | | | |
| Tiene dificultad para aprobar materias | <input type="radio"/> SI | <input type="radio"/> NO | | | |
| Es repitente | <input type="radio"/> SI | <input type="radio"/> NO | | | |
| Viene de pase | <input type="radio"/> SI | <input type="radio"/> NO | | | |
| | | | Volver | Finalizar | |

Imagen 61- Carga de datos (parte 2)

Al presionar *Finalizar*, los datos cargados se almacenan en la base de datos.

Si se presiona el botón *Cargar Inasistencias* de la Imagen 59, se verá otro menú (Imagen 62). Esta sección permite cargar la cantidad de inasistencias de los alumnos (uno por uno).

| MENU | | | | | Cerrar Sesión | | | | |
|---|---------------|--------------|----------------------|--------------|------------------|---------|---------------|--|--|
| Seleccione lo que desee realizar | | Nuevo Alumno | Cargar Inasistencias | Cargar Notas | Riesgo Academico | | | | |
| <table border="1" style="margin: auto;"> <thead> <tr> <th>Nombres</th> <th>Inasistencias</th> </tr> </thead> <tbody> <tr> <td colspan="2" style="background-color: #cccccc;"> </td> </tr> </tbody> </table> | | | | | | Nombres | Inasistencias | | |
| Nombres | Inasistencias | | | | | | | | |
| | | | | | | | | | |
| <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: auto;">Cargar.</div> | | | | | | | | | |

Imagen 62- Cargar inasistencias

Si se presiona el botón *Cargar Notas* de la Imagen 59, será posible cargar las notas de los alumnos por materia al finalizar el trimestre (Imagen 63)

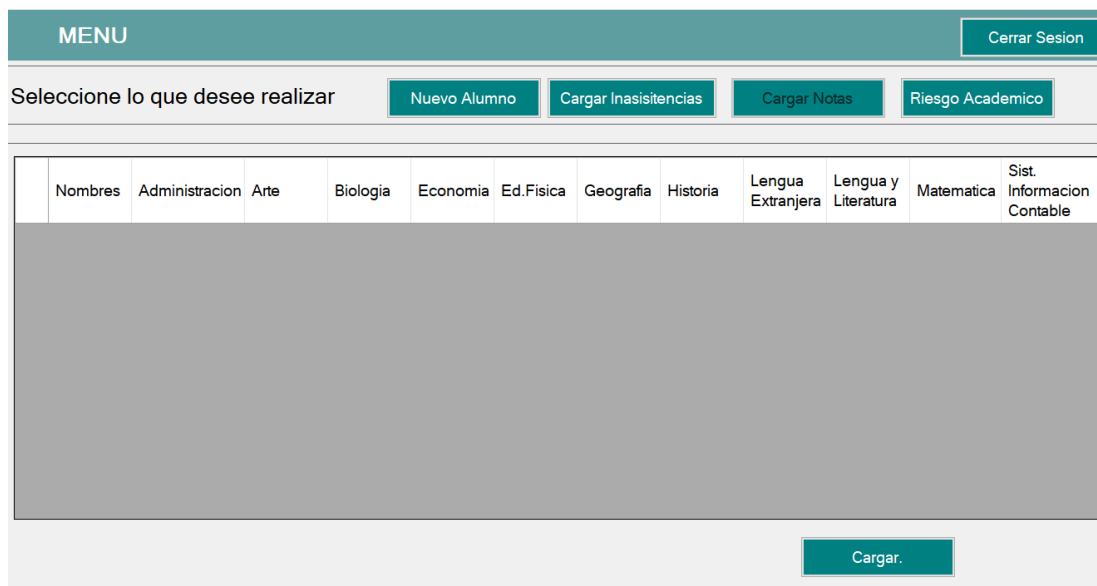


Imagen 63- Cargar notas

Las materias que se observan en la Imagen 63 corresponden al 3er año.

Si se presiona el botón *Riesgo Académico* (Imagen 63) se visualiza el riesgo académico (Imagen 64) que tiene cada alumno en base a:

- *Notas*: si tiene 5 o más materias desaprobadas, es decir, la mitad o más materias del total.
- *Inasistencias*: si tiene 10 o más inasistencias, es decir, la mitad o más inasistencias de las permitidas.
- *Otros factores*: es el resultado obtenido al aplicar el modelo de minería de datos.

| Nombre del Alumno | Notas | Inasistencias | Probabilidad de Riesgo por Factores Externos. | Información. |
|-------------------|-------|---------------|---|--------------|
| Estefanía | | | | VER |
| Juan | | | | VER |
| Camila | | | | VER |

Referencia.
Alumnos con riesgo

Imagen 64-Visualización del riesgo académico

El sistema, tiene un elemento importante al que denominamos *alarma*. Llamamos así al resaltado de color rojo que se visualiza en la interfaz del sistema (Imagen 64), cuando algún alumno se encuentra en situación de riesgo.

Además, se visualiza una planilla o grilla con cinco columnas (la Imagen 64): una columna con el *Nombre del alumno*, una columna denominada *Calificaciones*, otra columna cuyo nombre es *Inasistencias*, una columna llamada *Probabilidad de riesgo por Factores externos* y por ultimo una columna denominada *Información*. Cuando un alumno esté en situación de riesgo escolar se resalta la fila del alumno y la/s columna/s por las cuales se considera que está en riesgo.

En otras palabras, si un estudiante X se encuentra en riesgo académico debido a sus calificaciones, es decir, si obtiene notas bajas en la mitad o más de la mitad del total de materias que cursa, se resalta la columna *Calificaciones*. Si se encuentra en riesgo por inasistencias, es decir, que falta durante el trimestre 10 veces o más de 10 veces (ya que con 20 inasistencias quedaría libre), se resalta la columna *Inasistencias*. Si tiene probabilidad de estar en riesgo por otros factores, es decir, el resultado de haber aplicado el modelo de minería de datos al alumno, se resalta la columna *Probabilidad de riesgo por Factores externos*.

En caso de encontrarse en riesgo por las tres causas (notas, inasistencias y otros factores), se resaltan las tres columnas. En caso de no encontrarse en riesgo por notas, o por inasistencias u otros factores, la alarma no reporta ningún aviso, es decir, no resalta ninguna columna.

Para conocer con más detalle la situación del alumno, se presiona el botón VER de la columna *información* de la Imagen 64. Allí se podrá ver toda la información disponible: calificaciones, inasistencias e información personal de un alumno determinado.

Por ejemplo, para la alumna Estefania (este prototipo muestra solo el nombre del alumno, pero el sistema final mostrará el nombre completo y/o DNI que permita identificarlo) el sistema indica riesgo por inasistencia y por factores externos (Imagen 64). Podemos verificar esto, ya que, en la Imagen 65 se observa que Estefania tiene más de 10 inasistencias y además se observa que la alumna no vive con ambos padres, realiza actividades extraescolares, le cuenta aprobar las materias, viene de pase, entre otros.

| MENU | | | | | | | | | | | Cerrar Sesión | | | | | |
|------------------|---|------------------------------------|-----------------------------|--------------|---------------|---|-------------------|--------------------------|-------------|------------|------------------|---------------|---|----|---|--|
| Estefania | | | | | | | | | | | | | | | | |
| Notas: | Administraci | Arte | Biologia | Economia | Ed_Fisica | Geografia | Historia | Lengua_Ext | Lengua_Lite | Matematica | Sist_Inf_Cor | | | | | |
| ▶ | 9 | 8 | 9 | 6 | 9 | 7 | 1 | 9 | 5 | 9 | 10 | | | | | |
| * | | | | | | | | | | | | | | | | |
| Inasistencias: | <table border="1"> <thead> <tr> <th>Inasistencia:</th> </tr> </thead> <tbody> <tr> <td>▶</td> <td>11</td> </tr> <tr> <td>*</td> <td></td> </tr> </tbody> </table> | | | | | | | | | | | Inasistencia: | ▶ | 11 | * | |
| Inasistencia: | | | | | | | | | | | | | | | | |
| ▶ | 11 | | | | | | | | | | | | | | | |
| * | | | | | | | | | | | | | | | | |
| Otros: | Trabaja | Realiza actividades extraescolares | Le cuestan algunas materias | Es repitente | Viene de pase | Dedicacion mayor a dos horas al estudio | Posee plan o beca | Vive a mas de 10 cuadras | Movilidad | Tutor | El tutor trabaja | | | | | |
| ▶ | NO | SI | SI | NO | SI | SI | NO | SI | Colectivo | Padre | SI | | | | | |
| * | | | | | | | | | | | | | | | | |
| Cerrar | | | | | | | | | | | | | | | | |

Imagen 65-Visualización de datos personales y académicos de un alumno en particular

Anexo II. Alarma del sistema- Código.

Creación del archivo .CSV con datos almacenados en la base de datos

```
1. List<string> adminna = new List<string>();
2.   StringBuilder csvcontent = new StringBuilder();
3.   csvcontent.AppendLine("Edad;Vive Con Ambos Padres;Mas de dos
4.   hnos;Mayor de sus hnos;Nacionalidad Argentina;Vive Zona
5.   Rural;Trabaja Alumno;Actividad Extraescolar;Dificultad en
6.   Materia;Repitente;Pase;Dedica mas de 2hs
7.   diarias;Plan/Beca;Vive mas de 10
8.   cuadras;Movilidad;Tutor;Trabaja Tutor");
9.
10.   System.Data.DataTable dtl = new System.Data.DataTable();
11.   foreach (DataGridViewRow row1 in dataGridView1.Rows)
12.   {
13.       conexion.Open();
14.       SqlCommand consultall = new SqlCommand("SELECT Edad,
Vive_con_ambos_padres, Mas_de_dos_hnos, Mayor_hnos,
Nacionalidad_Arg, Vive_ZonaRural, Trabaja_Alumno,
Act_Extraescolar, Dificultad_Materias, Repitente, Vino_Pase,
Dedica_mas_de_Dos_Hs, Plan_Beca, Vive_mas_10_Cuadras, Movilidad,
Tutor, Trabaja_Tutor FROM Alumnos where nombre=@nombre",
conexion);
15.       consultall.Parameters.AddWithValue("@nombre",
Convert.ToString(row1.Cells["Nombre"].Value));
16.       SqlDataReader
reader1 = consultall.ExecuteReader();
17.       {
18.
19.           dtl = new System.Data.DataTable();
20.           if (reader1.HasRows)
21.           {
22.               SqlDataAdapter
daa = new SqlDataAdapter(consultall);
23.               reader1.Dispose();
24.               daa.Fill(dtl);
25.           }
26.       }
27.
28.       foreach (DataRow dr in dtl.Rows)
29.       {
30.           adminna.Add(dr["Edad"].ToString());
31.           adminna.Add(dr["Vive_con_ambos_padres"].ToString());
32.           adminna.Add(dr["Mas_de_dos_hnos"].ToString());
33.           adminna.Add(dr["Mayor_hnos"].ToString());
34.           adminna.Add(dr["Nacionalidad_Arg"].ToString());
35.           adminna.Add(dr["Vive_ZonaRural"].ToString());
36.           adminna.Add(dr["Trabaja_Alumno"].ToString());
37.           adminna.Add(dr["Act_Extraescolar"].ToString());
38.           adminna.Add(dr["Dificultad_Materias"].ToString());
39.           adminna.Add(dr["Repitente"].ToString());
40.           adminna.Add(dr["Vino_Pase"].ToString());
41.           adminna.Add(dr["Dedica_mas_de_Dos_Hs"].ToString());
```

```

35.     adminna.Add(dr["Plan_Beca"].ToString());
        adminna.Add(dr["Vive_mas_10_Cuadras"].ToString());
36.     adminna.Add(dr["Movilidad"].ToString());
37.     adminna.Add(dr["Tutor"].ToString()); adminna.Add(dr["Trabaj
a_Tutor"].ToString());
38.     }
39.     conexion.Close();
40.
41.     string csvpath = @ "D: \ Prueba1.csv" ;
42.     cadena cadena = "" ;
43.     string sub = "" ;
44.     for ( int x = 0 ; x < adminna . LongCount ( ) ; x ++ )
45.     {
46.         cadena v1 = Convertir . ToString ( adminna [ x ] ) ;
47.         if ( ( x != 0 ) && ( x % 17 == 0 ) )
48.         {
49.             sub = Convertir . ToString ( adminna [ x ] ) ;
50.             cadena coso = sub + ";" ;
51.             cadena = cadena + " \ n " + coso;
52.         }
53.         else
54.         {
55.             cadena = cadena + v1 + ";" ;
56.         }
57.         csvcontent . AppendLine ( cadena ) ;
58.         File . AppendAllText ( csvpath,Convert . ToString ( csvcont
ent ) ) ;
59.     }

```

Código para invocar el modelo de minería de datos generado.

```

1. ProcessStartInfo Info = new ProcessStartInfo("rutaRM");
2. Info.Arguments = @"Tesis/Proceso/PrediceNuevas";
3. Process p = Process.Start(Info);

```

Código para mostrar la alarma de riesgo académico por factores externos (resultado de la predicción del modelo de minería de datos).

```

1. //Tomando datos del Excel generado por el sistema.
2. Ruta del Excel en "Path1"

3. System.Data.DataTable tabla = new System.Data.DataTable();
4.     SLDocument s2 = new SLDocument(Path1);
5.
6.     int ii = 0;
7.
8.     while (!string.IsNullOrEmpty(s2.GetCellValueAsSt
ring(fila, 20)))
9.     {
10.         string riesgo = s2.GetCellValueAsString(fila, 20);

```

```
11.
12.             if (riesgo == "SI")
13.             {
14.                 dataGridView1.Rows[ii].Cells["Column15"].Style.BackColor = Color.Red;
15.             }
16.             else { dataGridView1.Rows[ii].Cells["Column15"].Style.BackColor = Color.White; }
17.             fila = fila + 1;
18.             ii = ii + 1; }
```

Código para mostrar el riesgo académico por Inasistencia.

```
1. //Consulta SQL:

SqlCommand consulta = new SqlCommand("select b.Administracion,
b.Arte, b.Biologia, b.Economia, b.Ed_Fisica, b.Geografia,
b.Historia, b.Lengua_Extranjera, b.Lengua_Literatura, b.Matematica,
b.Sist_Inf_Contable from AsignaturaAlumno b inner join Alumnos a on
b.Id_Alumno = a.Id_alumno where a.nombre = @nombre ",
conexion);

consulta.Parameters.AddWithValue("@nombre",
Convert.ToString(row.Cells["Nombre"].Value)); row.Cells["Nombre"].Value);

2. foreach (DataRow dr in dtt.Rows)
3.     {
4.         adminn.Add(dr["Cantidad_Inasistencia"].ToString());
5.     }
6.
7.     for (int x = 0; x < adminn.LongCount(); x++)
8.     {
9.
10.         int inasistencias = Convert.ToInt32(adminn[x]
11. );
12.         if (inasistencias >= 10)
13.         {
14.             dataGridView1.Rows[contadorfila].Cells["Column14"].Style.BackColor = Color.Red;
15.         }
16.         else
17.         {
18.             dataGridView1.Rows[contadorfila].Cells["Column14"].Style.BackColor = Color.White;
19.         }
20.         contadorfila = contadorfila + 1;
21.     }
```

Código para mostrar el riesgo académico por Notas.

```
1. //Consulta SQL:
```

```
2. SqlCommand consulta1 = new SqlCommand("select
Cantidad_Inasistencia from Inasistencia b inner join Alumnos a
on b.Id_Alumno=a.Id_alumno where a.nombre = @nombre ",
conexion);
3. consulta1.Parameters.AddWithValue("@nombre",
Convert.ToString(row1.Cells["Nombre"].Value));

4. foreach (DataRow dr in dt.Rows)
5.     {
6.         admin.Add(dr["Administracion"].ToString());
7.         admin.Add(dr["Arte"].ToString());
8.         admin.Add(dr["Biologia"].ToString());
9.         admin.Add(dr["Economia"].ToString());
10.        admin.Add(dr["Ed_Fisica"].ToString());
11.        admin.Add(dr["Geografia"].ToString());
12.        admin.Add(dr["Historia"].ToString());
13.        admin.Add(dr["Lengua_Extranjera"].ToString());
14.        admin.Add(dr["Lengua_Literatura"].ToString());
15.        admin.Add(dr["Matematica"].ToString());
16.        admin.Add(dr["Sist_Inf_Contable"].ToString());
17.    }
18.    int contadore = 0;
19.    for (int x = 0; x < admin.LongCount(); x++)
20.    {
21.        int nota = Convert.ToInt32(admin[x]);
22.        if (nota < 6)
23.        {
24.            contadore = contadore + 1;
25.        }
26.    }
27.
28.
29.    if (contadore > 3)
30.        dataGridView1.Rows[fil].Cells["Column1
31.3"].Style.BackColor = Color.Red;
32.    }
33.    else
34.    {
35.        dataGridView1.Rows[fil].Cells["Column13"].Style.BackColor = Color.White;}
36.    }
```

Anexo III. Instalación del sistema de alarma.

Para instalar y hacer uso del sistema de alarma se deberá realizar un conjunto de pasos detallados a continuación.

Pasos para instalar y usar el sistema de alarma.

- 1- Crear una carpeta vacía denominada *Sistema* en disco C:\.
- 2- Ejecutar e instalar el programa RapidMiner en la computadora, sin cambiar la ubicación que se crea por defecto (C:\Program Files\RapidMiner\RapidMiner Studio).
- 3- Una vez instalado, el programa RapidMiner se iniciará como en la Imagen 66.

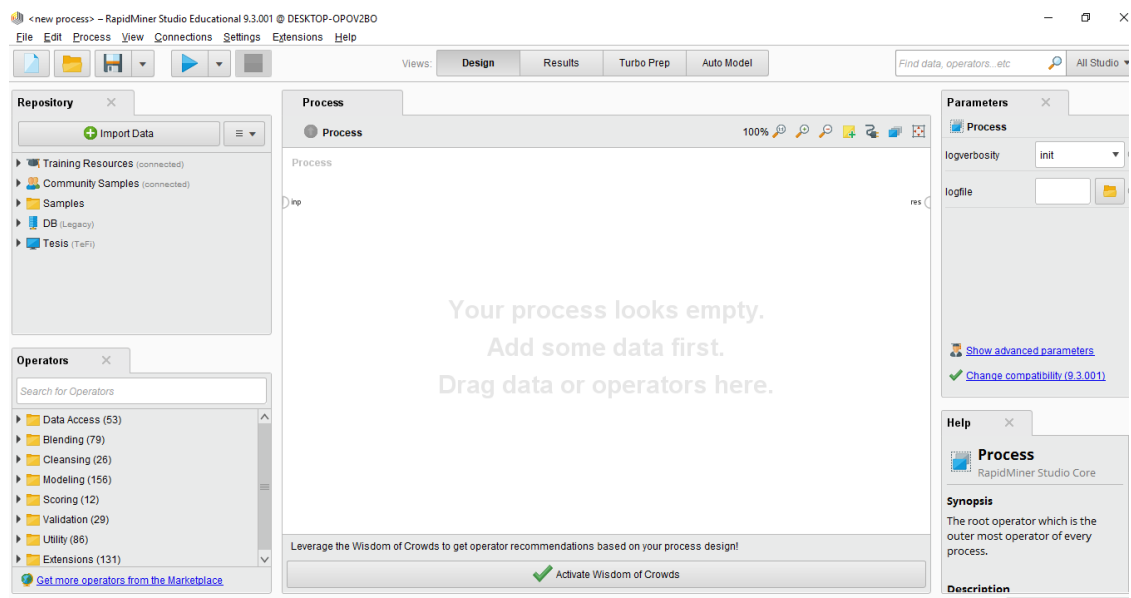


Imagen 66- Inicio de RapidMiner

- 4- Cerrar el programa RapidMiner.
- 5- Abrir el siguiente directorio: “C:\Users\Usuario\.RapidMiner\repositories”
- 6- Crear una nueva carpeta (repositorio) llamada *Tesis*.
- 7- Abrir la carpeta *Tesis* y crear dos subcarpetas denominadas *Proceso* y *Modelo*.
- 8- Copiar y pegar los archivos *Modelo.md*, *Modelo.ioo* y *Modelo.properties* en la subcarpeta *Modelo*.
- 9- Copiar y pegar los archivos *PrediceNuevas.rmp* y *PrediceNuevas.properties* en la subcarpeta *Proceso*.
- 10- Ejecutar el instalador de SQL Server.
 1. Hacer click en *Nueva instalación independiente de SQL Server*.



Imagen 67-Instalando SQL Server. Paso 1

2. Click en *Aceptar*.

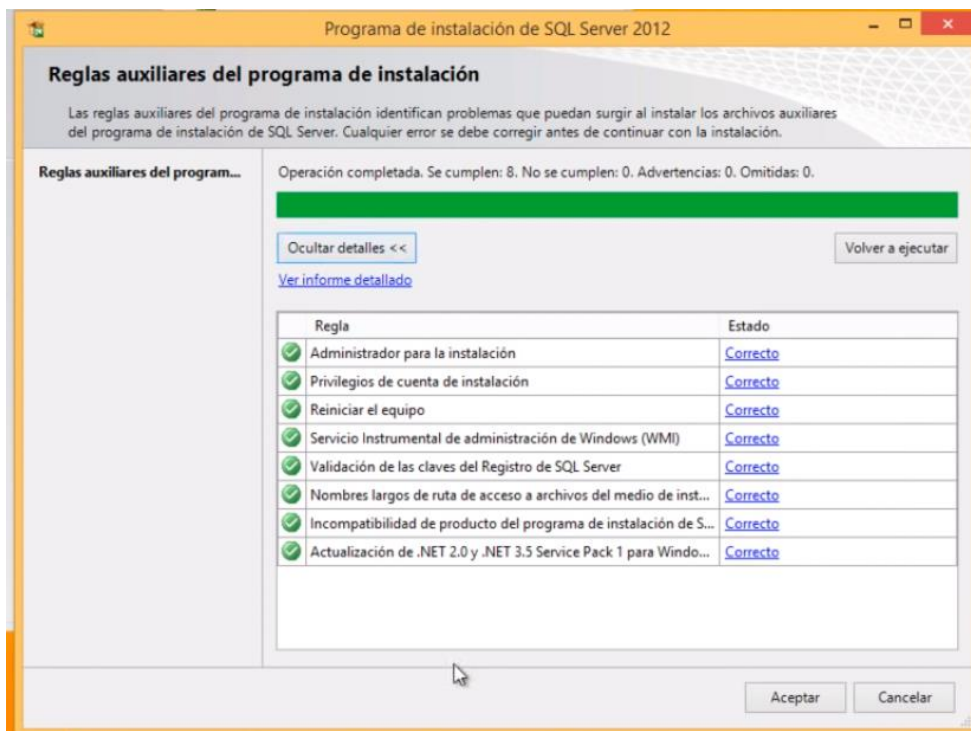


Imagen 68-Instalando SQL Server. Paso 2

3. Aceptar términos y condiciones de SQL Server y presionar *Siguiente*.

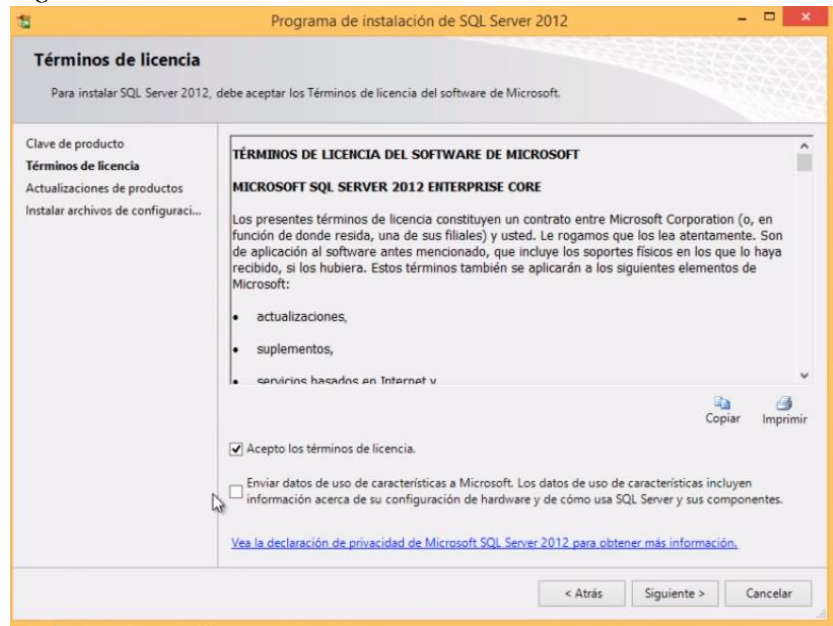


Imagen 69-Instalando SQL Server. Paso 3

4. Luego, presionar *Siguiente*.

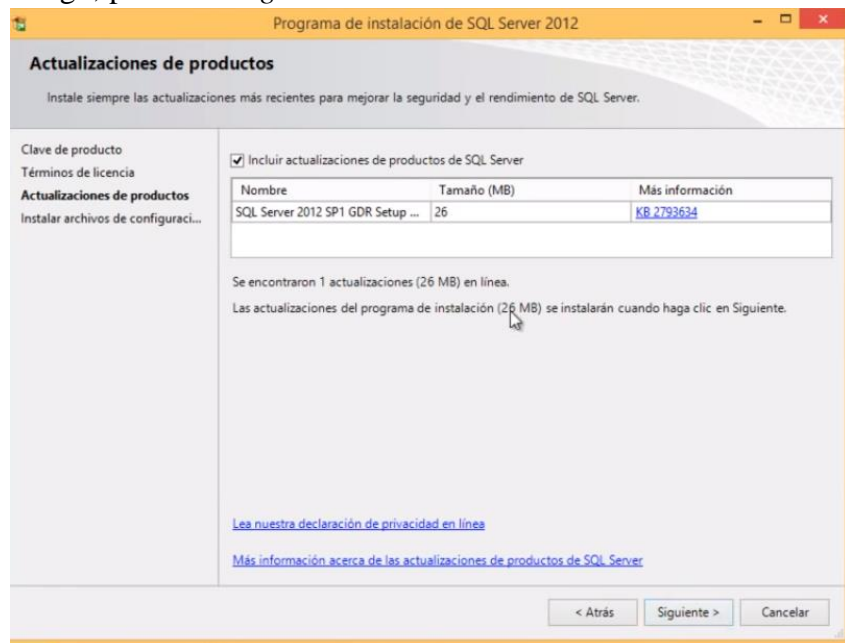


Imagen 70-Instalando SQL Server. Paso 4

5. Esperar unos minutos y presionar *Instalar*.

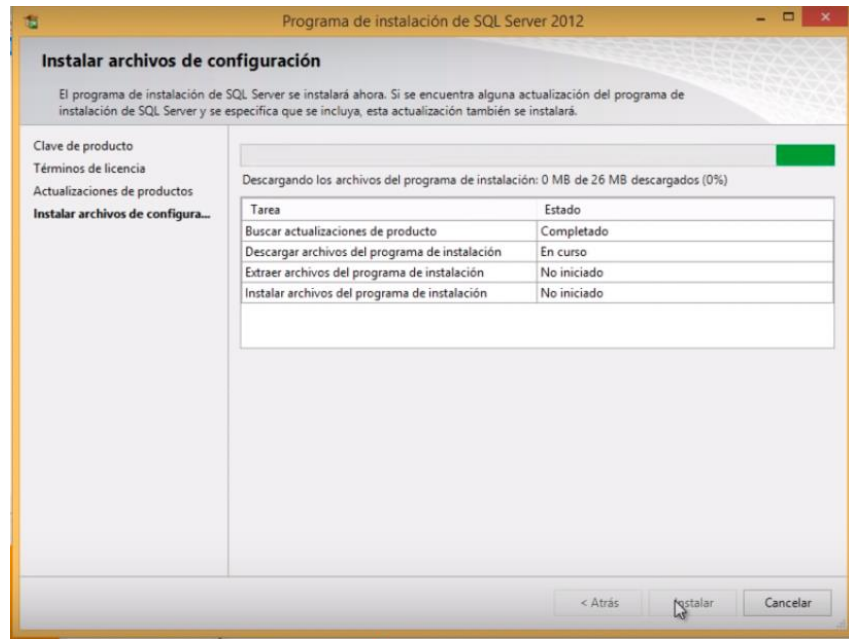


Imagen 71-Instalando SQL Server. Paso 5

6. Luego, presionar *Siguiente*.

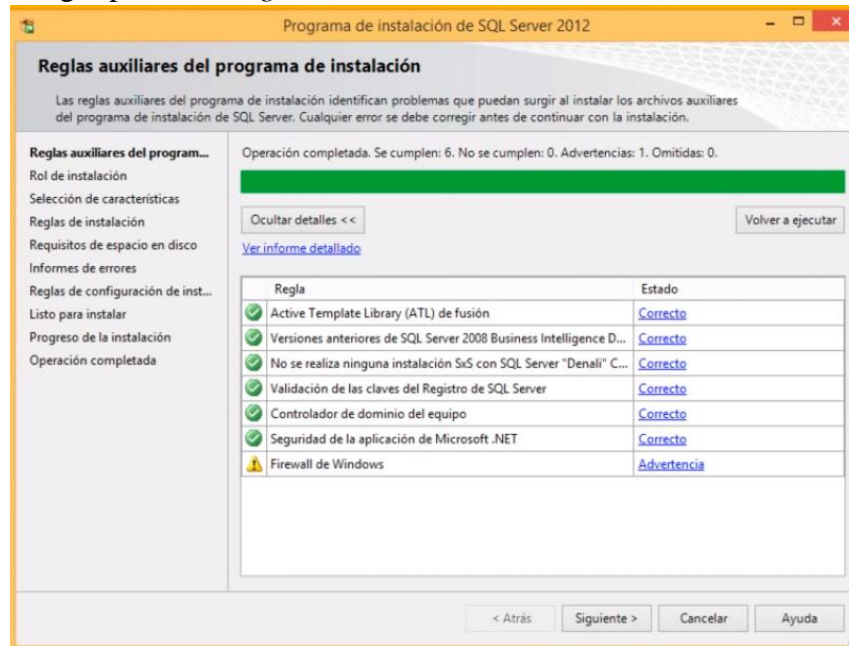


Imagen 72-Instalando SQL Server. Paso 6

7. Presionar *Siguiente*.

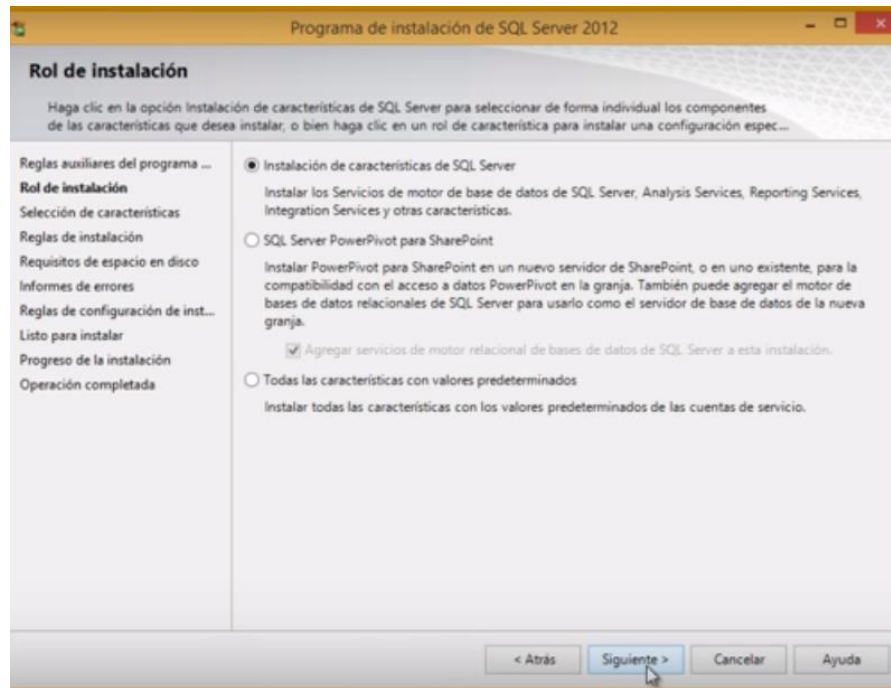


Imagen 73-Instalando SQL Server. Paso 7

8. Presionar *Siguiete*

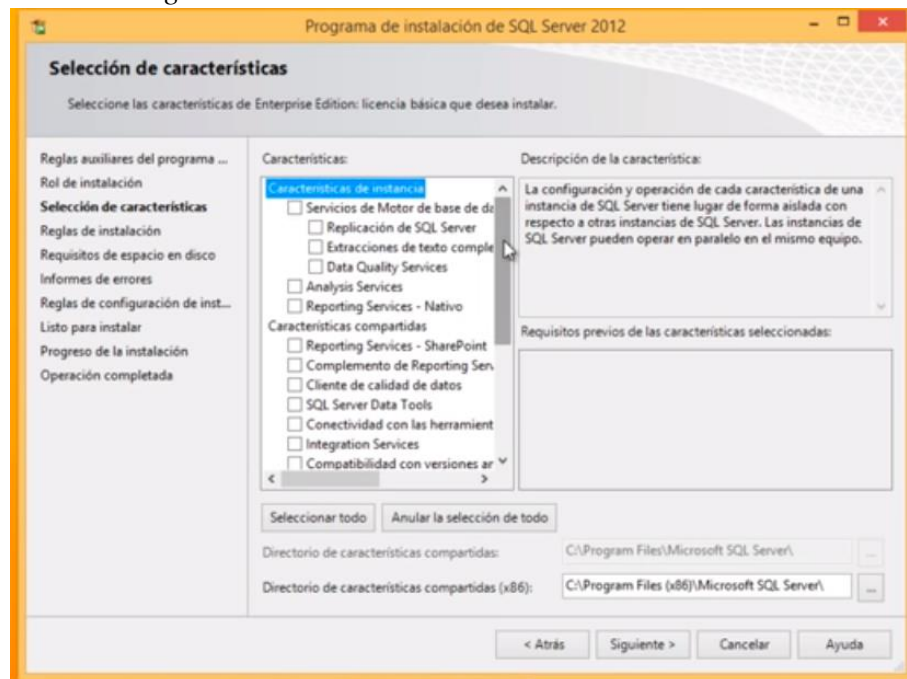


Imagen 74-Instalando SQL Server. Paso 8

9. Presionar *Siguiete*

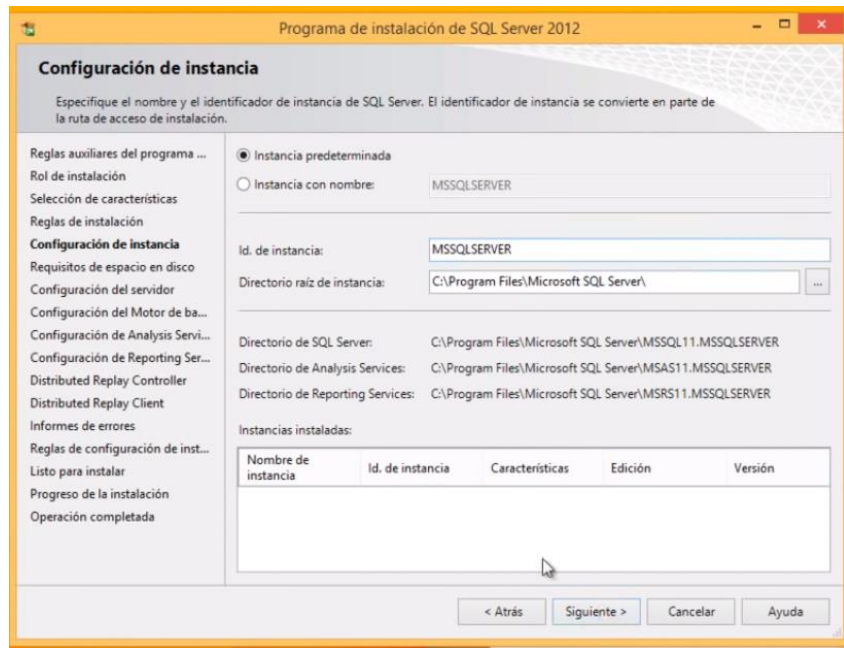


Imagen 75-Instalando SQL Server. Paso 9

10. Presionar *Siguiente*

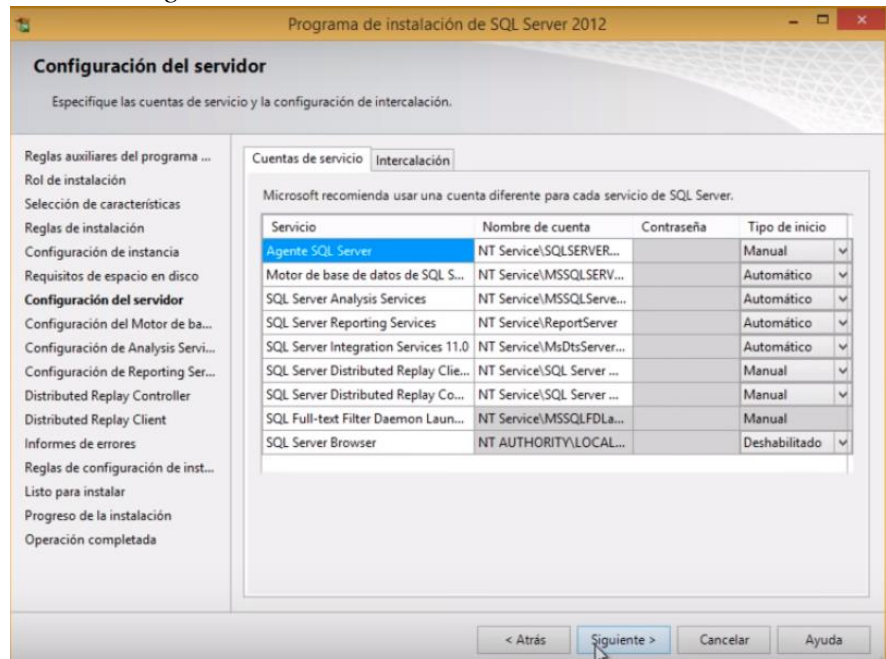


Imagen 76-Instalando SQL Server. Paso 10

11. Presionar *Siguiente*

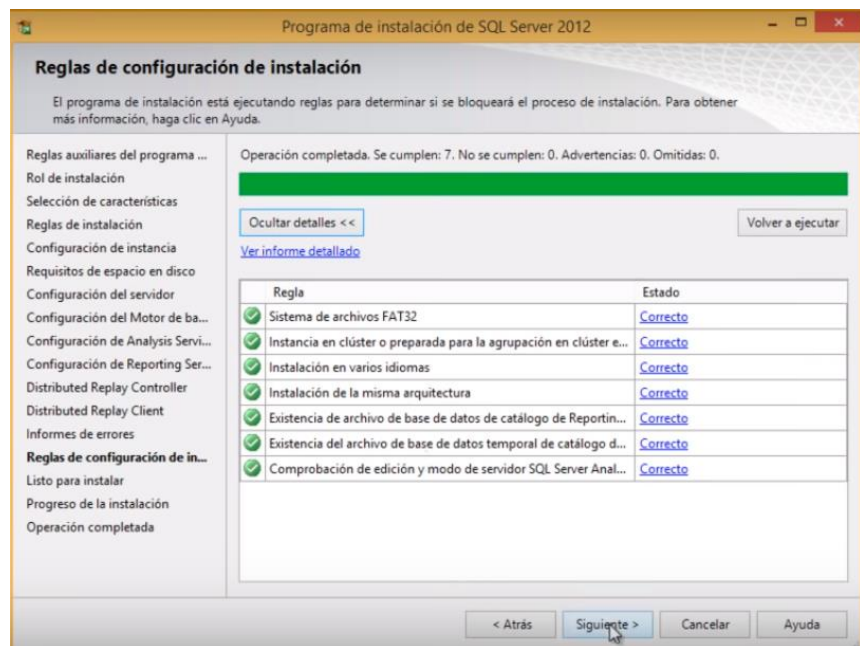


Imagen 77-Instalando SQL Server. Paso 11

12. Esperar que se instale el programa. Finalmente, aparecerá lo de Imagen 78. Presionar *Cerrar*.

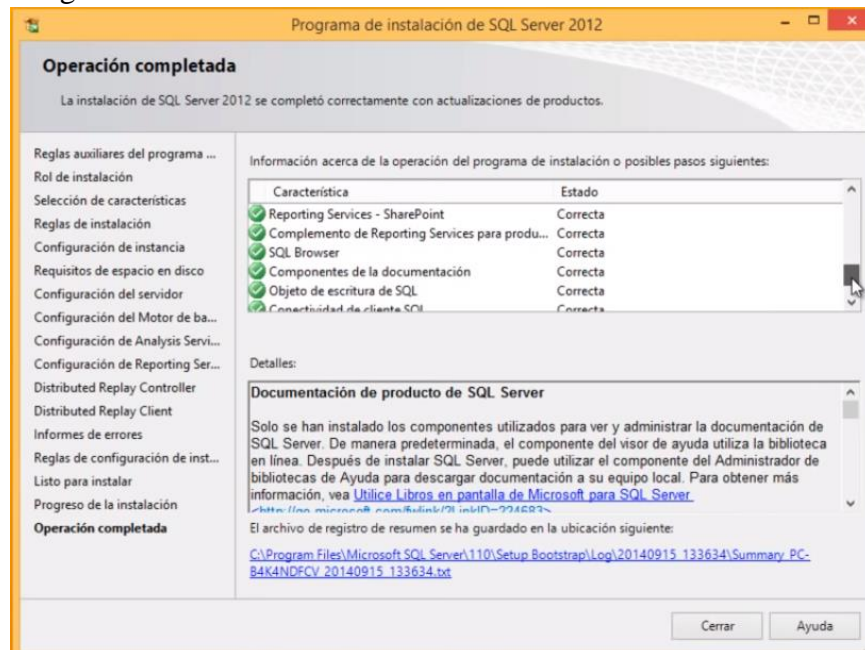


Imagen 78-Instalando SQL Server. Paso 12

- 11- Abrir SQL Server y copiar el nombre del servidor.

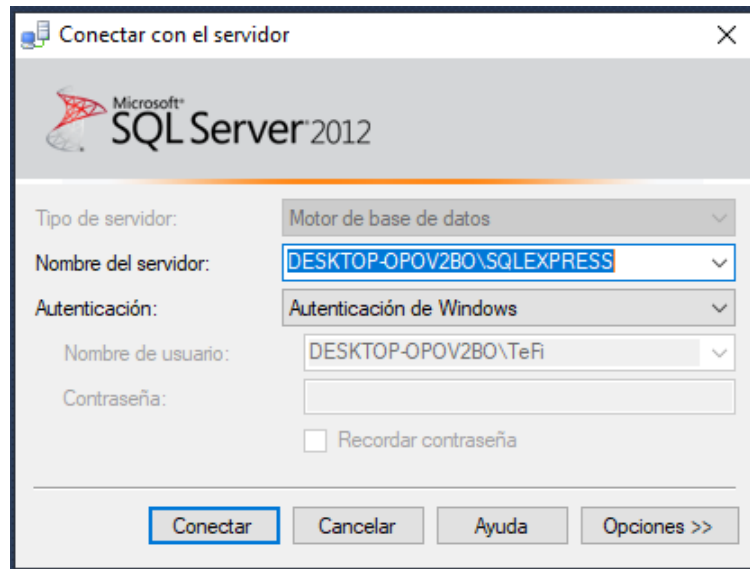


Imagen 79-Copiando nombre del servidor

- 12- Ejecutar el instalador del sistema de alarma llamado *Ap_escuela* y presionar *Siguiente*.
- 13- Ir a la carpeta en donde se instaló el programa y buscar el archivo *Ap_escuela.exe*. Luego, click derecho en él y *Editar*.

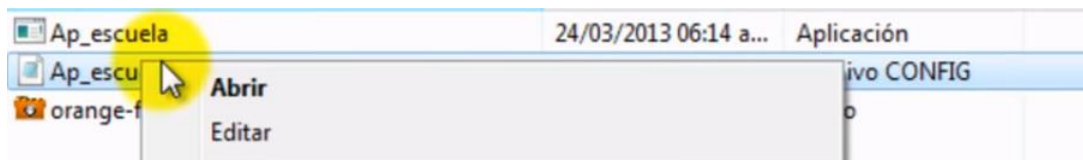


Imagen 80-Abriendo archivo CONFIG

- 14- Se abrirá el siguiente archivo.

```
<?xml version="1.0" encoding="utf8" ?>
<configuracion>
  <configSections>
  </configSections>
  <connectionString>
    <add name="Ap_escuela.Properties.Setting.Escuela.ConnectionString"
      connectionString="Data Source=DESKTOP-OPOV2BO\SQLEXPRESS database=Tesis; integrated security=true"
      providerName="System.Data.SqlClient" />
  </connectionStrings>
</configuracion>
```

Imagen 81- Modificando Data Source..

- 15- Copiar el nombre del servidor extraído en el paso 11 en la línea “*Data source*” del archivo y presionar *Guardar*.
- 16- En el escritorio, se verá el acceso directo creado llamado *Ap_escuela*. Hacer doble click en él.

Una vez realizados todos estos pasos, el sistema de alarma estará disponible para su uso.