



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**MODELO PREVENTIVO DE MOROSIDAD TEMPRANA DE CLIENTES HFC
MEDIANTE CLASIFICACION BAYESIANA**

**TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTION DE
OPERACIONES**

EDUARDO ANDRES BRUNA PAEZ

PROFESOR GUIA

RICHARD WEBER HAAS

MIEMBROS DE LA COMISION

JAIME MIRANDA PINO

SEBASTIAN MALDONADO ALARCÓN

JORGE RAMOS SEGURA

SANTIAGO DE CHILE

Octubre, 2012

RESUMEN

Este proyecto, formó parte del plan de mitigación de clientes morosos implementado por una empresa de telecomunicaciones en el año 2009, dónde una de las líneas de acción estaba enfocada en disminuir la morosidad de los nuevos clientes en su primera factura. Se propuso un modelo que clasifique a los nuevos clientes en pagadores y no pagadores, permitiendo con ello, reorientar recursos y realizar acciones preventivas de morosidad de manera más focalizada y efectiva, las cuales estaban siendo hasta ese momento aplicadas al universo total de nuevos clientes.

El éxito de este proyecto requería el compromiso del área de cobranzas, usuaria del modelo, por esto, la elección del modelo consideró aspectos tanto teóricos como prácticos. La simpleza y la fácil ejecución del modelo, idealmente programable, eran variables deseables y decidoras del éxito de éste. Se ejecutó basado en Clasificación Bayesiana por su forma simple de utilizar y presentar una confiabilidad aceptable [17]. Los métodos de inducción bayesiana han demostrado ser una clase de algoritmos tan competitivos como los métodos árboles de decisión y redes neuronales [29], que permiten implementar algoritmos en *Wolfram Mathematica 6.0* cuyas sentencias de programación son muy similares al lenguaje C++, permitiendo su programación posterior.

El modelo se confeccionó con 42.087 clientes, contratantes entre el 15 de julio y el 15 de octubre del 2009, de estos, 40.087 formaron la base de aprendizaje para calibrar y realizar los cálculos de probabilidades, y 2.000 como base de testing. Los clientes a clasificar en pagadores y no pagadores correspondieron a 9.328, quienes ingresaron a la empresa entre el 16 de Octubre y el 15 de Noviembre del 2009. En base a establecer una probabilidad a priori, se plantea en valores estimados de corte probabilístico basado en la experiencia o un método de clasificación [19], por una probabilidad de corte de 0,4 para clasificar a un cliente pagador por la necesidad que enfrenta la empresa de mejorar el % de clientes morosos en su primera boleta y a la vez reducir los costos actuales involucrados en la cobranza.

El modelo clasificó a 2.910 clientes como no pagadores, 2.179 de ellos efectivamente no cancelaron su boleta al vencimiento (75%). Este resultado es compatible con el esperado en el análisis teórico, debido a la eficiencia de los estimadores estimados, como también, el tamaño de los set de datos empleados que en forma teórica respalda los resultados obtenidos con un 80% de confiabilidad.

Como trabajo futuro, resulta interesante evaluar el aporte del modelo en la reducción real del porcentaje de clientes morosos al permitir reorientar recursos y focalizarse en ciertos clientes, como a su vez, actualizar la base de aprendizaje de tal forma ir considerando la actualidad al modelo.

DEDICATORIA

A mi padre, Eduardo, quién en vida me enseñó la importancia de tener hábito en todo lo que uno emprende.

A mi madre, Silvia, quién me ha enseñado a ver la vida de manera positiva, inculcándome la perseverancia como una herramienta para lograr los objetivos que uno se fija.

A mi pareja, Yanela, quién me ha acompañado día a día y me ha brindado el apoyo necesario durante estos años de estudio.

AGRADECIMIENTOS

A Jorge González, hoy integrante de nuestra familia, quién ha sido una persona muy generosa en conocimientos y traspaso de experiencia tanto de docencia como empresarial en la aplicación de modelos matemáticos para la toma de decisión empresarial.

Mis agradecimientos también al departamento de ingeniería industrial, en especial a Julie Lagos y a Richard Weber.

Tabla de contenido

1. DESCRIPCIÓN GENERAL EMPRESA PATROCINANTE	2
1.1. Tecnología HFC.....	5
2. JUSTIFICACIÓN DEL PROYECTO	7
3. METODOLOGÍA DEL PROYECTO	8
3.1. Método de clasificación y software matemático	9
3.1.1. Elección Método de Clasificación.....	9
3.1.1.1 Clasificadores y Comparaciones.....	11
3.1.1.2 Resumen y Justificación	15
3.1.2. Sistema y Clasificación Bayesiana.....	16
3.1.3. Redes Bayesianas Con Distribuciones de Probabilidad Condicionales	21
3.1.4. Estructura de Dependencia Bayesiano.....	23
3.1.5. Independencia Condicionada.....	25
3.1.6. Análisis de Independencia Condicional.....	27
3.1.7. Software de Programación	28
3.2. Bases de Aprendizaje, Testing y de Nuevos Clientes.....	29
3.3. Programación Modelo de Clasificación Bayesiana	43
3.3.1. Definición del Directorio de Origen de las Bases de Datos.....	43
3.3.2. Importación de Bases de Datos o Matrices.....	44
3.3.3. Codificación Bases de Aprendizaje y de Nuevos Clientes.....	48
3.3.4. Cálculo de la Probabilidad Bayesiana.....	58
3.3.5. Clasificación de Nuevos Clientes y Exportación de Resultados.....	69
4. RESULTADOS	73
5. CONCLUSIONES	76
6. BIBLIOGRAFÍA	78
7. ANEXO I: CALCULO DE PROBABILIDADES	82
7.1. Acontecimientos y sus operaciones lógicas.....	82
7.1.1. Acontecimientos	82
7.1.2. Operaciones lógicas de acontecimiento.....	83

7.2. Probabilidad en un conjunto finito	84
7.3. Equiprobabilidad y enumeración	86
7.3.1. Probabilidad Uniforme.....	86
7.3.2. Enumeración.....	87
7.4. Probabilidad condicional e independencia	88
7.5. Teorema De Bayes	93
7.6. Ejemplo de Clasificación Bayesiana	96

INTRODUCCION

En el último tiempo, las empresas de telecomunicaciones al ver que los segmentos corporativos, como también, el segmento de las grandes, medianas y pequeñas empresas están prácticamente “servidos”, y a su vez, fidelizados, empezaron a orientar su negocio al segmento residencial, y en especial al que se reconoce hoy como “mercado masivo”, es decir, aquellos hogares pertenecientes a la clase media – media baja, sobre todo en este último segmento, el cual está hoy en día accediendo al crédito y que está demandando productos y servicios que no son necesariamente de primera necesidad.

Dado lo anterior, se han desarrollado ofertas “paquetizadas” de servicios de televisión, internet y telefonía a precios altamente atractivos, dando la oportunidad de contratar sin mayores trámites con un modelo de instalación altamente efectivo y rápido, llegando incluso a tener una fuerza de venta acompañada con instaladores.

Estas ofertas trajeron consigo una variable a la que estas empresas no estaban muy acostumbradas, la morosidad e incobrabilidad. Tuvieron que reorientar sus áreas de recaudación y cobranzas e invertir en acciones constantes y masivas tanto de “recordatorio” como de cobro propiamente tal, que por cierto, han tenido resultados pero no los esperados, es más, enfrentan quejas de clientes molestos por los constantes llamados, mensajes de texto y emailing recordándole que su cuenta está por vencer o ha vencido.

Las interrogantes que ahora enfrentan estas empresas pasan por... ¿cómo seguir adquiriendo clientes para rentabilizar la red ya construida?, y a su vez, ¿cómo saber el comportamiento que tendrán en su primer pago?, ¿cómo reorientar los recursos y acciones en aquellos clientes que realmente no pagaran y así no molestar a quienes son buenos pagadores?... la búsqueda de respuestas a estas interrogantes dieron origen al presente proyecto, el cual busca a través del diseño y programación de un modelo de clasificación, permitir a la empresa, de manera eficiente, económica, rápida, simple de entender y bajo una lógica matemática, determinar cuáles de sus nuevos clientes se presumen como morosos en el pago de su primera factura al vencimiento de ésta.

1. DESCRIPCIÓN GENERAL EMPRESA PATROCINANTE

Por motivos de confidencialidad, en el presente proyecto, el nombre de la empresa será omitido. Se hará referencia a ésta indistintamente como empresa patrocinante o simplemente como la empresa, a si mismo, en esta sección, se describirá sólo de manera general los aspectos base necesarios para entender su negocio y tecnología a través de la cual entrega sus servicios.

La empresa dentro de su estrategia de penetración del mercado residencial en Chile, en Mayo del 2008, inició la oferta en el segmento masivo de servicios de televisión, internet y telefonía mediante la construcción de una red digital de tecnología denominada como HFC.

Actualmente la empresa está presente en las principales comunas de Santiago, como también , en regiones, en ciudades tales como Concepción, Talcahuano, Rancagua, Viña del Mar, Valparaíso, La Serena y Coquimbo.

Su oferta está compuesta por 12 planes base, pudiendo el cliente contratar televisión, internet ó telefonía por si solo, o bien, una combinación de estos, llegando a tener acceso si lo desea a los tres servicios en conjunto. Además, el cliente puede contratar de manera adicional, bolsas de telefonía, las cuales consisten en minutos adicionales de telefonía a los que ya trae el servicio base, bolsas de Internet, es decir, contratar megas adicionales, como también, decodificadores adicionales de tal forma de independizar cada uno de los televisores en el caso que su plan base contenga televisión. El rango del precio de lista de los planes base va desde los \$14.500 a los \$42.000, IVA incluido.

La base instalada de clientes a la fecha supera los 100.000 abonados, dónde el 31% de ellos posee los tres servicios instalados en sus hogares.

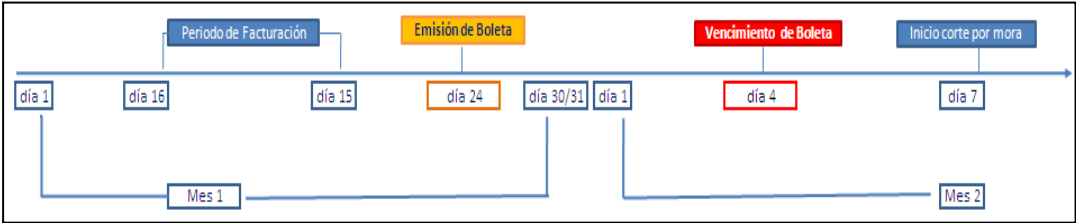
La empresa, utiliza dos canales de ventas, el canal de ventas Indirecto y el canal de ventas remoto. El canal de ventas indirecto tiene relación con la venta presencial, es decir, vendedores en terreno acompañados por instaladores que ofrecen el servicio puerta a puerta en los sectores con factibilidad técnica, por tal motivo, a través de este canal, el cliente puede contratar y contar con el servicio de manera prácticamente inmediata. Por otro lado, el canal de ventas remoto, es aquel dónde la venta se realiza vía telefónica, ya sea bajo modalidad outbound (llamada de salida) ó inbound (cliente llama). El canal de ventas remoto es menos inmediato, dado que posterior a la venta, ésta pasa a un proceso de certificación y es en esta validación dónde se acuerda con el cliente una fecha y hora de instalación del servicio.

Los segmentos económicos sobre los cuales la empresa focaliza su oferta está comprendido entre el segmento C2 y el segmento D. Esto obedece a que por un lado, en los segmentos económicos de mayores ingresos la penetración del servicio es alta, están altamente fidelizados y no acostubran a cambiarse por precio, y por otro, en Chile dada su actualidad económica, son los sectores económicos medios a medios bajos dónde se ha notado que están adquiriendo nuevos bienes y servicios que no necesariamente son de primera necesidad.

En lo que es el cobro del servicio utilizado, la empresa cuenta actualmente con un solo ciclo de facturación, el cual está compuesto por un periodo de facturación, que abarca desde el día 16 del mes en curso al día 15 del mes siguiente, por una emisión de boletas, la cual se realiza los días 24 de cada mes, día en el cual los clientes ya cuentan con su boleta publicada en la página web de la empresa, o bien, comienza la distribución hacia sus hogares vía correo privado. Un vencimiento que ocurre los días 4 de cada mes, por tanto, el cliente para no ser clasificado por la empresa como moroso tiene desde el día 24 del mes anterior al 4 del mes siguiente para realizar el pago de su servicio. Los lugares de pago, son los tradicionales, vale decir, Sencillito, Servipag, Cajas Vecinas de Banco Estado, entre otros. Por último, el ciclo de facturación

considera lo que se denomina como fecha de corte, que corresponde al día 7 de cada mes, y es ahí dónde se inicia el proceso de corte del servicio para todos los clientes que hasta esa fecha aún poseen su boleta impaga. Adicionalmente, es necesario destacar, que la incobrabilidad la empresa declara al día 90, por tanto, todo cliente que lleva más de 90 días sin haber pagado su última factura emitida, se declaran en baja por morosidad y se inicia la gestión de retiro de equipos desde sus hogares. La figura 1 resume de manera gráfica el ciclo de facturación anteriormente descrito.

Fig. 1. Ciclo facturación servicio.



Fuente: Empresa.

1.1. Tecnología HFC

La empresa es proveedor de televisión, internet y telefonía a través de la tecnología que se conoce bajo el nombre de HFC, de sus siglas del inglés "Hybrid Fibre Coaxial" ("Híbrido de Fibra y Coaxial"). En Telecomunicaciones, es un término que define una red que incorpora tanto fibra óptica como cable coaxial para crear una red de banda ancha. Esta tecnología permite el acceso a internet de banda ancha y telefonía utilizando las redes de televisión por cable, conocidas como CATV existentes. Se puede dividir la topología en dos partes. La primera consiste en conectar al abonado (cliente) por medio de cable coaxial a un nodo zonal y posteriormente interconectar los nodos zonales con fibra óptica. Esta tecnología comienza a implementarse a través de operadores de CATV, que además de brindar el servicio de televisión por cable anexaron transportar por el mismo medio la señal de internet de banda ancha y telefonía.

A través del uso de cada una de estas tecnologías, la red es capaz de aprovecharse de los beneficios y minimizar el impacto de las limitaciones inherentes a cada una.

La fibra óptica proporciona la ventaja de cubrir distancias razonablemente largas con un mínimo de amplificación y regeneración de la señal. Sin embargo, debido a la naturaleza de esta tecnología, el costo y tamaño de los multiplexores/desmultiplexores ópticos, rara vez se utiliza para conectar los nodos directamente a los clientes. En lugar de eso, la fibra óptica termina en un elemento de la red llamado "puerta de enlace" (o gateway), el cual contiene, al menos, un transformador óptico (normalmente, dos) que permite la transición de la señal a la red de cable coaxial. El cable coaxial proporciona una capacidad

de ancho de banda considerable, mientras que también permite que la señal se extraiga y se inserte con una mínima interferencia a cualquier cliente o equipo. Las limitaciones de este sistema son que a veces la señal necesita ser amplificada y además es susceptible a interferencias externas.

2. JUSTIFICACIÓN DEL PROYECTO

La empresa, ante la necesidad de ganar rápidamente cuota de mercado frente a sus competidores en el segmento residencial, en especial en aquellos hogares pertenecientes a la clase media y media baja, ha establecido como estrategia llegar con servicios a precios altamente competitivos y sin barreras discriminatorias de ingreso (scoring comercial), en el sentido de no solicitar antecedentes que permitan determinar en parte, en un cliente, su capacidad de pago, ingreso económico, o bien, su condición de deuda actual con otros servicios.

Si bien es cierto estas acciones le han permitido captar mensualmente aproximadamente 15.000 nuevos clientes, la característica del segmento ha traído como consecuencia que la morosidad del servicio esté sobre los parámetros establecidos y esperados por la empresa, ante esto, se establecieron acciones de “cobranza preventiva” a todo nuevo cliente que ingresa a la empresa, no obstante, los resultados que se han logrado no son los esperados en comparación al costo involucrado de ejecutar estas acciones, y más aún, se incrementó la insatisfacción de los nuevos clientes dado que por un lado se sienten “molestados” por este tipo de acciones, y por otro, sienten que la empresa “desconfía” de ellos, sobre todo cuando en la práctica son buenos pagadores.

Dado el escenario anterior, la empresa a través de este proyecto, busca contar con un modelo matemático soportado en alguna tecnología que le permita clasificar mensualmente, de manera rápida y oportuna, todo nuevo cliente, identificando aquellos clientes que si pagarán su primera factura al vencimiento, y excluirlos de las acciones de “cobranza preventiva” y así no molestarlos, y orientar de mejor forma los esfuerzos y recursos en “los malos pagadores”

3. METODOLOGÍA DEL PROYECTO

La metodología de trabajo para el desarrollo de éste modelo preventivo de morosidad temprana consta de 7 fases, las cuales se desarrollarán a lo largo de este capítulo.

Esta metodología comienza con decidir qué modelo de clasificación utilizar para diferenciar a los futuros clientes en dos categorías, pagadores y no pagadores. Junto con la elección de la teoría de clasificación, se debe elegir el software matemático sobre el cual se ejecutará el modelo, recordar que será utilizado a nivel usuario por el área de cobranzas de la empresa, y por tanto, se desea programar por el área de informática de la empresa de tal forma que las acciones a ejecutar por esta área sean prácticamente carga de nuevos clientes y extracción de los clientes ya clasificados y así realizar las acciones pertinentes.

Realizado lo anterior, se determinan y analizan los clientes que serán parte de la base de aprendizaje del modelo, como a su vez, se determinan los clientes que formarán parte de la base de testing de éste, clientes de ambas bases que a priori se sabe cuál fue su comportamiento de pago de su primera factura cuando contrataron el servicio. Junto con estos clientes, se determinan los nuevos, es decir, los futuros clientes sobre los cuales se aplicará el modelo de clasificación. Definidos, se determina que campos se poseen en las bases de datos y que entregan información de los clientes de tal forma definir cuáles de ellas serán utilizadas como variables predictoras del modelo y cuál será la variable discriminante que permitirá al modelo clasificar a los clientes en pagadores y no pagadores de su servicio en su primera factura al vencimiento de ésta. Con los antecedentes anteriores ya determinados, se comienza con la fase de programación del modelo, para luego validarlo con la base de testing, continuando con la clasificación de los nuevos clientes en pagadores y no pagadores, y terminando esta metodología de trabajo cruzando la clasificación

realizada por el modelo para los nuevos clientes con el comportamiento real que tuvieron y de esta forma medir su efectividad. Esto en base a la clasificación bayesiana y lo obtenido por el modelo estimado.

La figura 2 grafica la metodología del proyecto.



Fig. 2 Metodología modelo preventivo de morosidad temprana. Fuente: Elaboración propia.

3.1. Método de clasificación y software matemático

3.1.1. Elección Método de Clasificación

Actualmente los métodos de clasificación o clasificadores cobran más fuerza debido a lo complejo de los datos a analizar, el cual es necesario abordar una gran cantidad de datos, a partir de esto, se hace más necesario mejores instrumentos de clasificación, pero este proceso es complejo dependiendo del contexto en el que estos se presenten, pueden tener diferentes significados y traer diferentes consecuencias como fraude financiero, anomalías cardíacas, accidentes de tránsito aéreo, intrusión en la red de datos, clasificación de clientes confiables en un banco, violación de seguridad informática, etc. [7]

A partir de lo anterior, se abordan distintos algoritmos de clasificación donde el área de la computación aborda este problema, puesto que existen pocos programas disponibles en el mercado que debido a la complejidad de algunos métodos se deben programar en esta tarea. Los modelos de algoritmos más conocidos para resolver este tipo de problemas son redes neuronales, árboles de decisión, reglas de decisión y redes bayesianas. Dentro de estos modelos, los árboles de decisión son los más utilizados en esta tarea [7], siendo los algoritmos ID3 y sus variantes los más conocidos (como C4.5, J4.8). En general, este tipo de algoritmos se caracterizan porque son entrenados en una fase de aprendizaje en la que aprenden a modelar el comportamiento de los datos/eventos. Para realizar dicho propósito, el conjunto de datos se divide en training set y test set.

- i) El training set es el conjunto de datos utilizado para modelar el comportamiento del clasificador.
- ii) El test set es el conjunto de datos no conocidos en el momento del entrenamiento, para medir la exactitud de predicción del algoritmo clasificador.

Determinar un clasificador eficiente, requiere de una serie de análisis previos y determinar la naturaleza de los datos e hipótesis propuestas, de esta forma no es una tarea fácil. Cada clasificador se caracteriza porque emplea una representación diferente de los datos. Encontrar una buena representación de estos, que calce con el problema a resolver, requiere de tiempo y de varios ensayos previos. En realidad, aún no existe una regla general acerca de qué métodos de clasificación son los más apropiados para cuales tipos de problemas.

Según estudios previos realizado por la Universidad de Chile en 2005, en el ámbito de clasificar datos, constituye un hecho esencial determinar características específicas que en sí que permitan establecer el mejor método

de clasificación. Se ha demostrado que ciertas partes del espacio de datos es mejor modelado por un método de clasificación en comparación a otros. El empleo de diferentes clasificadores puede proporcionar información complementaria importante sobre la representación de los datos. Esto ha originado que utilizar una combinación o ensamble de clasificadores sea una buena alternativa.

Para efectos del estudio, se basaran en tres métodos de clasificación como lo son: [25]

- i) Redes neuronales
- ii) Árboles de decisión
- iii) Redes bayesianas.

A través de estudios experimentales, muestran que no existe “el mejor método”, si no que más bien dependiendo de las características del espacio de los datos e hipótesis cual es el más apropiado, aunque puede ocuparse más de un método de clasificación es necesario determinar estas características y elementos diferenciadores basadas en las hipótesis del estudio a realizar. Adicionalmente, en este último tiempo varias investigaciones, bs métodos de inducción bayesiana han demostrado ser una clase de algoritmos tan competitivos como los métodos árboles de decisión y redes neuronales [24 y 25], y se están usando exitosamente en muchas aplicaciones relacionadas con la clasificación [4].

3.1.1.1 Clasificadores y Comparaciones

Un clasificador provee una función que mapea/clasifica un registro de datos dentro de una de las distintas clases predefinidas (Kohavi, 1996). Se requiere de varios ensayos previos para encontrar un método de clasificación que sea el mejor en las tareas o problemas a solucionar, y que se comporte también como

un buen ensamble de clasificadores. [18] Bajo esta premisa, existen ensambles de clasificadores que permiten combinar y determinar cuál (es) de ellos son los más óptimos. A través de un ensamble, permiten combinar de alguna forma las decisiones individuales de cada uno de ellos para clasificar nuevas instancias [12]. Existen varias razones que justifican el ensamble de clasificadores, entre ellos se tiene:

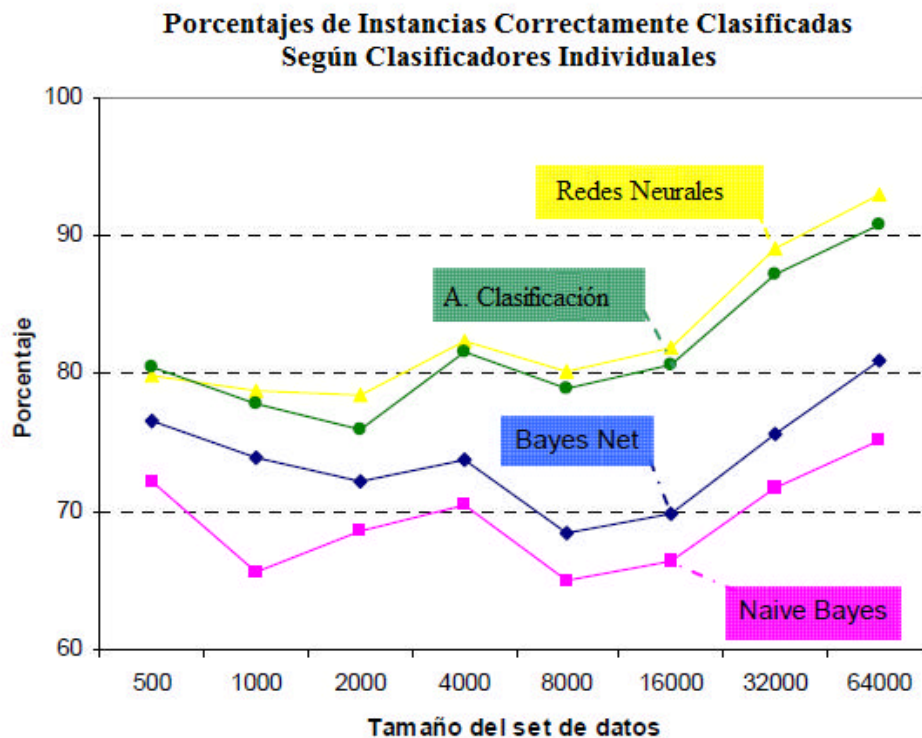
- Diferentes clasificadores utilizan diferentes técnicas y métodos de representación de los datos, lo que permite obtener resultados de clasificación con diferentes patrones de generalización.
- Los datos pueden no proveer suficiente información para elegir un único mejor clasificador debido a que el tamaño disponible en estos datos es pequeño en comparación al espacio de hipótesis [12]

A partir de algoritmos complejos, es posible determinar cuál de los clasificadores presentan mejores resultados con otros métodos de clasificación. De esta forma [12], plantea combinar clasificadores, el cual puede ser una buena alternativa cuando se pretende obtener un mejor desempeño que el reportado en la ejecución individual de cada clasificador. En este último tiempo, las redes bayesianas y el clasificador Naive Bayes (NB) han logrado captar la atención de los investigadores, pero a pesar de ello, aún no queda claro si estos algoritmos bayesianos logran mejorar significativamente su predicción al ser combinados con otros algoritmos de clasificación.

Para evaluar, que tan precisos constituyen estos clasificadores, [21], mediante el uso del algoritmo LRTree es considerado como un punto de referencia entre el desempeño de algoritmos híbridos nativos y algoritmo híbridos construidos con métodos de ensamble. Los clasificadores son evaluados en forma independiente sobre el set de datos experimental y de control, todos con diferente número de instancias. Los set de datos experimentales presentan la misma distribución con el fin de determinar el

impacto en el desempeño de los clasificadores cuando éstos procesan más evidencia.

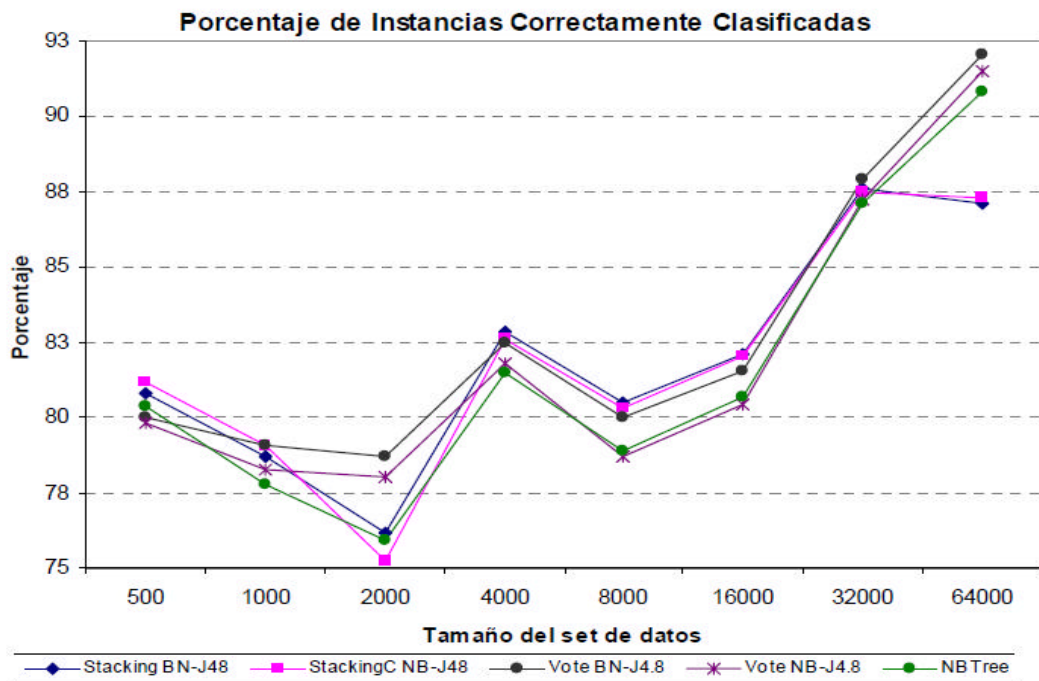
A partir de la evidencia muestral, los resultados obtenidos muestran que los algoritmos bayesianos mejoran significativamente la exactitud de predicción al ser ensamblados con el árbol de decisión J4.8. Ahora bien en términos individuales, si bien, la exactitud de predicción de los algoritmos bayesianos reportó ser menor en comparación al algoritmo J4.8 basado en los arboles de clasificación y de redes neurales, en los set de datos experimentales ésta puede ser mejorada si los parámetros de configuración de ejecución son afinados; tales como, el estimador a emplear, o algoritmo de búsqueda a utilizar. Es decir, por un lado incrementar el número de iteraciones que provee mayor precisión sobre los resultados obtenidos y obtener un estimador más eficiente, se refiere a tener por un lado una mayor cantidad de observaciones y menor variancia. También se puede concluir que la exactitud de predicción de un clasificador individual o de un ensamble puede ser mejorada con un set de evidencia bastante representativo del problema. [22]. La exactitud de un ensamble definitivamente puede ser mayor a la exactitud de predicción obtenida por los algoritmos individuales base. En la figura siguiente muestra la comparación individual de cada clasificador en base a la aplicación experimental basado en el algoritmo LRTree.



A partir del grafico anterior, se tienen buenos resultados a medida que aumenta el tamaño de los datos, Sin embargo, en los set de datos con mayor número de instancias, se puede mejorar el rendimiento tanto en la predicción y en clasificador, en este caso, según [22], al combinar los algoritmos bayesianos como BN y NB con los arboles de clasificación a través del algoritmo NBTree se logran mejores rendimientos hasta en un 15% el porcentaje de clasificación, es decir que al combinar el ensamble de clasificadores, es posible mejorar su exactitud. A partir de lo anterior, se desprende que en la medida que los algoritmos ensamblados procesan set de datos con mayor número de instancias, tienden a mejorar su exactitud de clasificación, así mismo, el tiempo de procesamiento también se incrementa.

[12], señala que se puede mejorar las instancias correctamente clasificadas, combinando los algoritmos bayesianos BN y NB con el árbol de clasificación. La principal ventaja corresponde en un 15% el porcentaje de clasificación al ser ensamblados con el árbol de clasificación. Sin embargo, la exactitud de

clasificación del ensamble no refleja ser mejor a la obtenida por los clasificadores individuales que los componen, en particular al desempeño logrado por el árbol de decisión J4.8. [17]



Finalmente, la reducción de la dimensión de los set de datos es un tema importante por realizar, así como el ensamble de algoritmos bayesianos con redes neuronales, lo cual es también soportado por lo métodos de ensamble el cual pueden mejorarse considerablemente estos métodos bayesianos de clasificación y proporcionar mejoras en los pronósticos.

3.1.1.2 Resumen y Justificación

A través de los antecedentes de los puntos anteriores, los resultados obtenidos muestran que los algoritmos bayesianos corresponden a una buena alternativa sobre otros métodos clasificación como las redes neuronales, obteniendo una eficiencia en la exactitud de la predicción sobre el 80% para

muestras grandes sobre 64 mil en el set de datos. Para propósitos de la presente investigación se logran sin problemas en el conjunto de datos para cumplir con las exigencias teóricas que respaldan la investigación. Además de una fácil implementación en la parte práctica a diferencia de otros métodos. Si bien, la exactitud de predicción de los algoritmos bayesianos reporta ser menor en comparación a otros métodos, ésta puede ser mejorada si los parámetros de configuración de ejecución son afinados; tales como, el estimador a emplear, o algoritmo de búsqueda a utilizar. También se puede concluir que la exactitud de predicción de un clasificador individual o de un ensamble puede ser mejorada con un set de evidencia bastante representativo del problema. La exactitud de un ensamble definitivamente puede ser mayor a la exactitud de predicción obtenida por los algoritmos individuales base. [19]

3.1.2. Sistema y Clasificación Bayesiana

El objetivo de un sistema Bayesiano es saber cuál es la hipótesis más probable entre varios conjuntos de datos, es decir, dado un conjunto de datos (que dividiremos en dos conjuntos de entrenamiento y de test) representados por pares <atributo, valor>, el problema consiste en encontrar una función $f(x)$ (llamada hipótesis) que clasifique dichos ejemplos. [6].

La idea principal en el sistema bayesiano consiste en usar el teorema de Bayes en cualquier problema de aprendizaje automático (en especial los de clasificación) es que podemos estimar las probabilidades a posteriori de cualquier hipótesis consistente con el conjunto de datos de entrenamiento para así escoger la hipótesis más probable.

Si $P(D)$ es la probabilidad a priori de los datos, $P(D|h)$ su probabilidad dada una hipótesis y se desea estimar $P(h|D)$, la probabilidad posterior de h dados los datos. Se puede plantear el siguiente teorema:

$$P(H_i | D) = \frac{P(D | H_i) P(H_i)}{\sum_j P(D | H_j) P(H_j)}$$

La hipótesis más probable o MAP (máximo a posteriori hipótesis)

$$H_{MAP} = \underset{H_i}{\operatorname{arg\,max}} \frac{P(D | H_i) P(H_i)}{\sum_j P(D | H_j) P(H_j)}$$

Según [4], el algoritmo Bayesiano tiene la ventaja que puede ser fácilmente implementado si se calculan todas las posibles hipótesis en la ecuación 1 y se selecciona la hipótesis de mayor probabilidad. En general si tiene un sistema de aprendizaje de lo general a lo específico (o al revés) que busca especializaciones más generales (o generalizaciones más específicas) se puede caracterizar asumiendo que las hipótesis más generales (o específicas) son más probables que otras.

Un aspecto relevante que determina un elemento diferenciador, es el principio básico que puede ser empleado para determinar de una forma a priori la clase en la cual puede ser clasificado un dato, según las probabilidades generadas por una serie de funciones de clasificación bayesiana, previamente definidas. Generalmente en otros métodos de clasificación no se disponen estas probabilidades, según la teoría bayesiana al no tener información disponible se opta por un máximo verosimilitud que en este caso corresponde a valor a priori definido por el usuario.

Tomando como referencia estos principios, ciertos autores como [25] y [27], han establecido funciones de evaluación Bayesiana, las cuales se basan en definir una función de probabilidad para cada una de las clases, que se puede establecer mediante un algoritmo de aplicación.

De otro lado, es importante señalar que los algoritmos basados en un enfoque bayesiano han sido usados en varias y diversas áreas de la ingeniería como:

- Aproximación de funciones de valores discretos, capaz de expresar hipótesis disyuntivas y robustas frente a factores de ruido [8].
- Clasificación estadística, aplicada al reconocimiento de imágenes [9].
- Identificación de fallos, basados en series de reconocimiento de patrones, generados por los valores históricos de las variables observadas [10].
- Identificación y pronóstico de cartera de clientes con morosidad de pagos basados en optimizar una futura cartera de clientes confiables. [11].

De esta forma, podemos señalar que los clasificadores estadísticos, desde un enfoque bayesiano, han sido ampliamente usados en la solución de diversas clases de problemas, principalmente por presentar una mayor robustez frente a factores no deseados como el ruido; aspecto que los convierte en un tema de gran interés para la investigación. [12].

Si bien, otros clasificadores basados en redes neuronales, máquinas de soporte vectorial, y otros, parecen obtener resultados ligeramente superiores, en condiciones de normalidad, es importante resaltar que una metodología empleada en método bayesiano en uso de algoritmos como el matlab, Mathematica o estadística es más fácil de implementar y robusta ante factores de ruido; además puede ser empleada para reducir parámetros acústicos y seleccionar los más efectivos [12,13,14,15]

Otro aspecto importante es la potencia del clasificador, que es una prueba estadística que depende de qué tan diferente en realidad es la media verdadera de la población del valor supuesto. La potencia es la probabilidad de rechazar la hipótesis nula cuando esta es falsa, es decir, es la probabilidad que cuando en realidad no se cumple la hipótesis nula que planteamos, el procedimiento estadístico nos da como resultado que se rechaza la hipótesis nula. En términos absolutos, se tiene una potencia adecuada sobre el valor 0.7, entre más cercano a uno, se tiene una mayor precisión.

Según [9], en este clasificador se obtiene una efectividad cercana al 85%, para muestras grandes, aspecto que lo convierte en un algoritmo confiable en el reconocimiento de una clase a priori, a la cual puede pertenece una característica. [23], señala que sobre una potencia del 75% se considera confiables los clasificadores.

Como se mencionó en capítulo anterior, encontrar un clasificador eficiente no es una tarea fácil. Cada clasificador se caracteriza porque emplea una representación diferente de los datos, en caso de los clasificadores bayesianos, se pueden lograr buenos resultados para mayores instancias y mejoras en la eficiencia de los estimadores en términos de determinar probabilidades a priori en los puntos de cortes distintos al valor 0,5.

A fin de tener mayor claridad en los clasificadores bayesianos, se describen los métodos de clasificación más importantes.

- i) Las Redes Bayesianas (BN) representan las dependencias que existen entre los atributos a través de una distribución de probabilidad condicional en un grafo dirigido acíclico (DAG) [26]. La dependencia entre dos atributos es descrita por la presencia de un arco entre ellos, y su influencia causal, por la dirección del arco. La independencia entre atributos se representa por la ausencia de un arco que conecte atributos particulares.
- ii) El clasificador Naive Bayes (NB) es un caso especial de una red bayesiana, en el que se asume que los atributos son condicionalmente independientes dado un atributo clase [27], de ahí que sea llamado también como “clasificador ingenuo”. Este clasificador es muy eficiente en ciertos dominios, debido a que es muy robusto ante atributos irrelevantes.

Tomando como referencia en este último clasificador comprende el objetivo del presente estudio a fin de determinar un porcentaje de instancias correctamente clasificadas y como elemento predictor.

De esta forma, en forma cuantitativa (y ésta es la gran aportación de los métodos bayesianos), da una medida probabilística de la importancia de esas variables en el problema (y por lo tanto una probabilidad explícita de las hipótesis que se formulan). Esta es quizá una de las diferencias fundamentales que ofrecen las redes bayesianas con respecto a otros métodos -como puedan ser los árboles de decisión y las redes neuronales-, que no dan una medida cuantitativa de esa clasificación.

En términos predictivos, los métodos bayesianos permiten tener en cuenta en la predicción de la hipótesis el conocimiento a priori o conocimiento del dominio en forma de probabilidades. El problema puede surgir al tener que estimar ese conocimiento estadístico sin disponer de datos suficientes y no tener una probabilidad a priori, el cual según el apartado anterior, se requieren grandes cantidades de datos a fin de mejorar la potencia y eficiencia de los estimadores, el cual constituye una dificultad sobre los otros métodos de clasificación (Redes Neuronales y Árboles de Clasificación). Esta dificultad ha sido estudiada por L. M. Bergasa [22], que analizaron los sesgos que se producen en los sujetos en la estimación subjetiva de las probabilidades de un suceso.

3.1.3. Redes Bayesianas Con Distribuciones de Probabilidad Condicionales

En relación a la teoría que sustenta esta investigación, las Redes Bayesianas, constituyen una alternativa con alta confiabilidad en la clasificación que puede pertenecer una característica y su fácil implementación. [17], sustenta además la importancia de la estructura de la independencia y la independencia condicionada, el cual es consecuencia directa de la estructura en la red bayesiana como la distribución normal multivalente. Otras propiedades como la independencia de las variables normales incorreladas o como la determinación de independencias condicionadas por los ceros en la inversa de la matriz de covarianzas, hacen más fácil el cálculo y el tratamiento local de las variables.

Por la definición de Red Bayesiana, se sabe que dicho modelo se especifica en función de un Diagrama de Flujo que presenta las variables del problema con sus relaciones de dependencia y un conjunto de distribuciones de probabilidad P . En el caso de las Redes Gaussianas, la distribución condicionada de cada variable, asociada a una distribución normal univariante.

Por lo tanto, para obtener la Red Bayesiana Gaussiana descrita a partir de la distribución condicionada de cada una de las variables, es necesario especificar los siguientes parámetros.

$$\mu_j$$

Donde μ_j es la media marginal de la variable X_j , para todo $j \in \{1, \dots, n\}$

$$\sigma_{jj}^2$$

Donde σ_{jj}^2 es la varianza condicionada de la variable X_j , para todo $j \in \{1, \dots, n\}$

$$\sigma_{jk}^2$$

Siendo β_{jk} , con $\beta_{jj} = 1$, el coeficiente de regresión de X_j en el modelo de regresión de X_j sobre X_k , $j, k \in \{1, \dots, n\}$

Con esta información, obtenida bien por observación e información de diversos estudios, bien por un análisis de datos previo, es posible definir una Red Bayesiana Gaussiana calculando la distribución conjunta normal multivalente como el producto de las densidades condicionadas.

A partir de lo anterior, se tienen los elementos para la construcción de una Red Bayesiana Gaussiana con la probabilidad conjunta. Se trata de especificar una distribución de probabilidad conjunta normal multivalente $N(\mu, S)$.

En este caso, los expertos han de especificar los valores que componen el vector de medias μ y la matriz de covarianzas S , bajo la condición de ser dicha matriz una matriz definida positiva. [24] Por consiguiente, para definir la Red Bayesiana Gaussiana que describe un problema, se deben fijar los elementos de:

?? Vector de medias, dado por μ_j para todas las variables X_j , para todo $j = 1, \dots, n$

Σ : Matriz de covarianzas simétrica definida positiva, donde σ_{jj} es la varianza de X_j y σ_{ij} es la covarianza entre X_i y X_j .

3.1.4. Estructura de Dependencia Bayesiano

Shachter [26] estudian la tratabilidad de los modelos gráficos Gaussianos en una estructura grafica concreta, el diagrama de influencia, describiendo una transformación de los parámetros condicionados $\{v_1, \dots, v_n\}$ y $\{\beta_{ij} \text{ con } j < i\}$ para obtener la matriz de covarianzas S definida positiva. Con esta finalidad, desarrollan un algoritmo que muestra la relación entre el modelo gráfico Gaussiano y la representación de la matriz de covarianzas para una distribución normal.

En los modelos gráficos Gaussianos, la matriz de covarianzas S es de gran interés, ya que recoge la estructura de dependencia de las variables del problema mediante la matriz inversa de S , como se verá en el siguiente Apartado. Algunos desarrollos teóricos de estos modelos no exigen que la matriz de covarianzas S sea definida positiva, por ello en esta Tesis el interés

se centra en el desarrollo de Shachter, et al. [26] que especifican una matriz definida positiva, que será necesaria para posteriores desarrollos en el modelo gráfico dado por una Red Bayesiana Gaussiana.

Apoyándose en resultados de Yule [30] y en la descomposición de Chan, H., Shachter, [9, 26] concluyen que la matriz de covarianzas S se puede descomponer como el producto de diversas matrices, siendo dicha matriz S definida positiva e invertible o regular. Con esta finalidad, se definen las siguientes matrices:

- D : matriz diagonal formada por las varianzas condicionadas v_i , siendo $v_i > 0$ para todo i .
- B : matriz triangular inferiormente formada por los coeficientes de regresión β_{ji} para todo $j < i$.

Como B es triangular inferiormente, B es invertible y por tanto, se puede definir la matriz $U = B^{-1}D^{-1}$. De esta forma, la matriz de covarianzas S se obtiene como:

$$S = U^{-1} D U^{-T}$$

Como la matriz U es regular, la transformación de S dada es congruente, y por tanto, si todas las varianzas condicionadas son distintas de 0, la matriz S es definida positiva. En caso de existir algún $v_i = 0$, entonces dicha matriz es semidefinida positiva.

Sea la matriz de precisión K , la matriz de covarianzas inversa dada por S^{-1} . Se sabe que si S es regular, entonces dicha matriz de precisión se puede obtener como:

$$K = S^{-1}$$

La matriz de precisión K es de gran interés en el estudio de los modelos gráficos Gaussianos, por reflejar relaciones de independencia entre las variables del problema. En el caso de las Redes Bayesianas Gaussianas dicha matriz permite obtener relaciones de **independencia condicionada** entre las variables.

A partir de los conceptos empleados, en forma práctica, para determinar la independencia de las variables predictoras es necesario establecer las siguientes proposiciones cuyas demostraciones matemáticas se citan a los autores correspondientes. [13]

3.1.5. Independencia Condicionada

Han sido muchos los autores que han trabajado con la distribución normal multivalente. En este Apartado, se presentan dos resultados conocidos de dicha distribución que relacionan la independencia condicionada con la matriz de covarianzas S . Las demostraciones de las proposiciones propuestas pueden verse en [21] o [2].

Como la matriz de covarianzas S está formada por varianzas en la diagonal, y por covarianzas en el resto de elementos, se puede determinar la independencia entre dos variables en función de los ceros de la matriz S . La siguiente proposición, refleja esta idea.

Proposición 1

Sea $X \sim N(\mu, S)$ tal que X se particiona en dos componentes X_1 y X_2 siendo μ y S .

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ y } \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

Entonces X_1 y X_2 **son independientes** si y solo si S_{12} es la matriz **0**. [18]
 Si la matriz de covarianzas S es regular, es decir existe su inversa, entonces esta proposición se puede enunciar también si y solo si $S_{12} = 0$. Con este resultado, a continuación se presenta una proposición que fija la independencia condicionada de las variables con la distribución normal multivariante, en función de los ceros que aparecen en la matriz de precisión $K = S^{-1}$ de dicha distribución. Esta proposición es consecuencia directa de la estructura de la distribución normal condicionada.

Proposición 2

Sea $X \sim N(\mu, S)$, donde S es una matriz regular (o invertible). Entonces, las variables X_i y X_j son condicionalmente independientes dado el resto de las variables de X , si y solo si el valor de la posición (i, j) de la matriz K es cero, es decir:

$$K_{ij} = 0$$

Donde $K = S^{-1}$ es la matriz de precisión de la distribución.

[18]

Por tanto, a partir de la matriz de covarianzas se pueden conocer las relaciones de independencia condicionada existentes entre las variables del problema, identificando los ceros en la matriz de precisión **K**.

A continuación, se presentan las relaciones de independencia condicionada entre las variables predictoras del problema del estudio, en relación a la base de aprendizaje en función de los resultados anteriormente expuestos.

3.1.6. Análisis de Independencia Condicional

A partir del estudio presente, se van a describir las relaciones de independencia condicionada existentes entre las siete variables predictoras y su discriminante $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$, el cual se deben primeramente estar en forma discreta definidas en base a la categorización establecida de acuerdo a la base de aprendizaje. Estas se describen en forma más detallada en el punto 3.2.

A partir de esto, se establece las relaciones de independencia en base a la Matriz de Covarianzas que describe la distribución conjunta normal multivariante. Para esto se utiliza el programa SPSS.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1	0	0	0	0	0	0	0
X_2	0	1	0	0	0	0	0	0
X_3	0	0	1	0	0	0	0	0
X_4	0	0	0	1	0	0	0	0
X_5	0	0	0	0	1	0	0	0
X_6	0	0	0	0	0	1	0	0
X_7	0	0	0	0	0	0	1	0
X_8	0	0	0	0	0	0	0	1

Con la matriz de covarianzas se observa independencia entre las variables X_1 y X_6 , X_1 y X_8 , X_2 y X_6 , X_2 y X_8 , X_3 y X_5 , X_3 y X_6 , X_3 y X_8 , X_4 y X_6 , X_4 y X_8 , X_5 y X_8 , X_6 y X_8 , X_7 y X_8 . De acuerdo a la proposición 1, todas las variables son

independientes si existe una matriz O , fila o columna. En este caso A^{-1} es la matriz O . Además esta matriz es no singular puesto que su determinante es distinto de cero, es decir que existe su inversa.

```

      ? ? ? ?      ?      ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
      ?      ? ? ? ? ? ? ? ? ? ? ? ? ?      ?      ? ? ? ? ? ? ? ?
? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
? ? ? ? ?      ?      ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
      ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?

```

3.1.7. Software de Programación

Como se ha comentado a lo largo de este trabajo, hay consideraciones relevantes realizadas por la empresa, una de ellas es la posibilidad de poder programar el modelo de tal forma que el área usuaria sólo tenga que ingresar bases de clientes a clasificar, y extraer bases de clientes clasificados en pagadores y no pagadores para realizar posteriormente acciones sobre estos últimos. De ahí que el algoritmo de clasificación bayesiana es una de las partes relevantes a desarrollar en este proyecto ya que será la base y guía para la programación del modelo por el área de informática de la empresa.

Considerando lo anterior, el algoritmo de cálculo del modelo de clasificación bayesiana fue programado utilizando Wolfram Mathematica 6.0, software matemático que tiene la particularidad de asemejarse al lenguaje de programación C++ facilitando con ello el desarrollo a los programadores de la empresa.

3.2. Bases de Aprendizaje, Testing y de Nuevos Clientes

Se ha determinado en conjunto con la empresa que los clientes a formar parte de la base de aprendizaje y testing del modelo serán todos los clientes instalados entre el 15 de julio y el 15 de octubre del 2009, equivalente a 42.087 clientes. De este total, 40.087 clientes se utilizarán para calibrar y realizar los cálculos de probabilidades que exigen la clasificación Bayesiana, y otra parte, 2.000 clientes para validar los resultados del modelo calibrado.

Los nuevos clientes a clasificar en pagadores y no pagadores, y sobre los cuales se validará la efectividad del modelo, en el sentido que se evaluará clasificación versus real comportamiento, corresponderán a todas las instalaciones del servicio realizadas entre el 16 de octubre y el 15 de Noviembre de 2009, equivalente a 9.328 clientes.

Las bases de clientes de la empresa están compuestas de siete campos, los cuales describen a cada cliente. Adicionalmente, apoyados del área de cobranzas de la empresa para las bases de aprendizaje y testing se logró obtener un octavo campo que tiene relación cual fue el comportamiento de pago de cada uno de estos clientes de su primera factura cuando ingresaron a la empresa.

Se cuenta entonces con siete variables que describen a un individuo y una que indica su comportamiento.

Considerando la experiencia de quienes mes a mes están enfocados en disminuir la morosidad de clientes, se les consultó su apreciación con respecto a las variables y su efecto en la morosidad del cliente, el resultado es que las consideran altamente relevantes a cada una de ellas y que encuentran adecuado que en este modelo sean consideradas, se dará a conocer en la descripción de cada variable los argumentos entregados. Teniendo en cuenta la

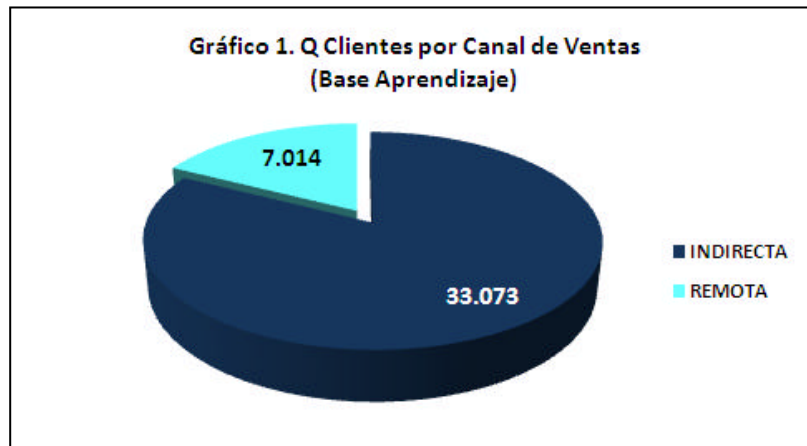
opinión de la empresa, el que clasificador Naive Bayes asume independencia condicional de clase, aún cuando además, se realizó un análisis de independencia de las variables, previa discretización de las variables de tipo cuantitativa en cualitativas (ver 3.1.5), en este proyecto se consideraron las siete variables como predictoras, y como variable discriminante, para las bases de aprendizaje y testing, el estado de la morosidad del cliente, octava variable. La justificación en el uso de las siete variables predictoras corresponde a variables estratégicas a considerar por la empresa al momento de captar un producto a ofertar y su importancia en el efecto de la variable discriminante.

A continuación se describe cada una de las variables, predictoras y discriminante, indicándose las modalidades que la constituyen y la cantidad de clientes por modalidad tanto de la base de aprendizaje, testing como la de nuevos clientes.

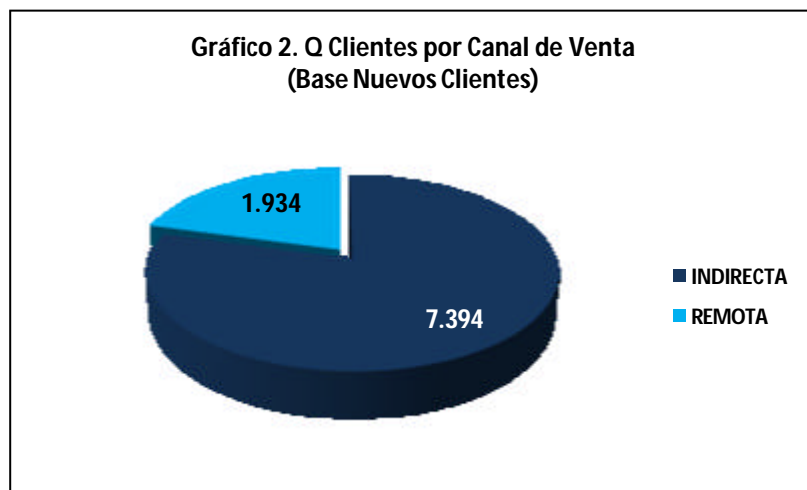
- a) **Canal de venta:** Variable cualitativa nominal. Relacionada con el tipo de canal por el cual el cliente contrató el servicio, vale decir, si la contratación fue mediante ejecutivos de terreno (canal indirecto), o bien, vía ejecutivos de call center (canal remoto).

La justificación de considerar esta variable por la empresa, pasa por el hecho de que ambos canales aun cuando se rigen por las mismas condiciones comerciales, sus procesos de certificación de ventas son distintos, siendo el canal remoto, aquel que posee mayores filtros o validaciones de certificación de la venta, es decir, velar que el cliente efectivamente entiende en detalle lo que está contratando. En cambio, el canal de venta indirecto, acostumbra a vender bajo una modalidad denominada como inmediata, es decir, la venta e instalación del servicio ocurre en el momento de la contratación. Dado lo anterior, se intuye que la morosidad puede verse afectada por la presencia o ausencia de filtros

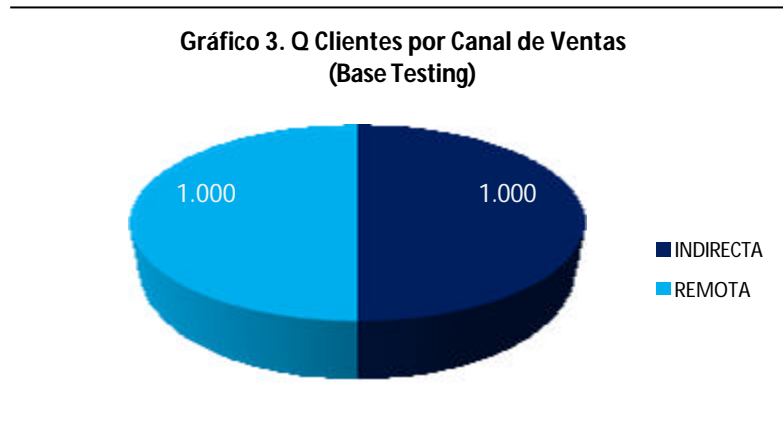
de certificación. Las modalidades que componen a esta variable son dos, terreno y remota.



Fuente: Empresa.



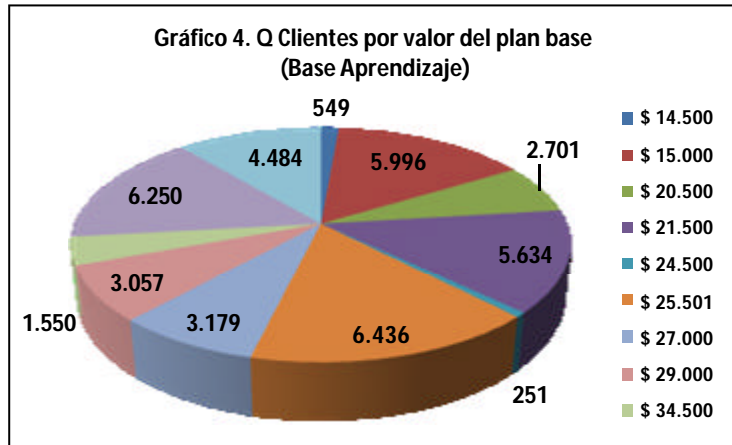
Fuente: Empresa.



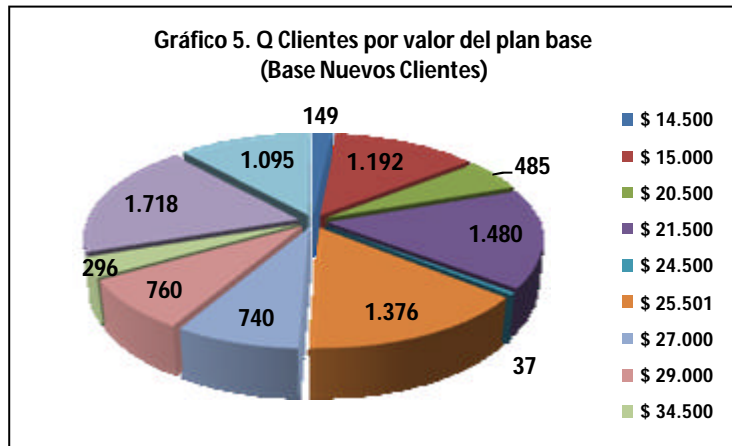
Fuente: Empresa.

- b) **Valor del plan base:** Variable cuantitativa discreta. Consiste en el valor o precio de lista del servicio base contratado por el cliente, se denomina base dado que no considera el costo incurrido por el cliente en la contratación de servicios adicionales, tales como, decodificadores, canales premium, bolsas de minutos de telefonía y/o internet. Se ha determinado su presencia en el modelo, dado que se considera que el monto mensual a pagar por un cliente condiciona su comportamiento de pago del servicio.

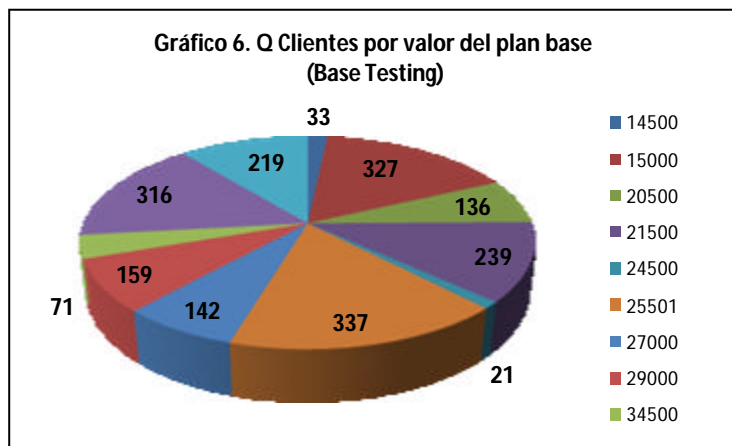
Está compuesta de 11 modalidades, siendo estas: \$14.500, \$15.000, \$20.500, \$21.500, \$24.500, \$25.501, \$27.000, \$29.000, \$34.500, \$37.500, \$42.000.



Fuente: Empresa



Fuente: Empresa

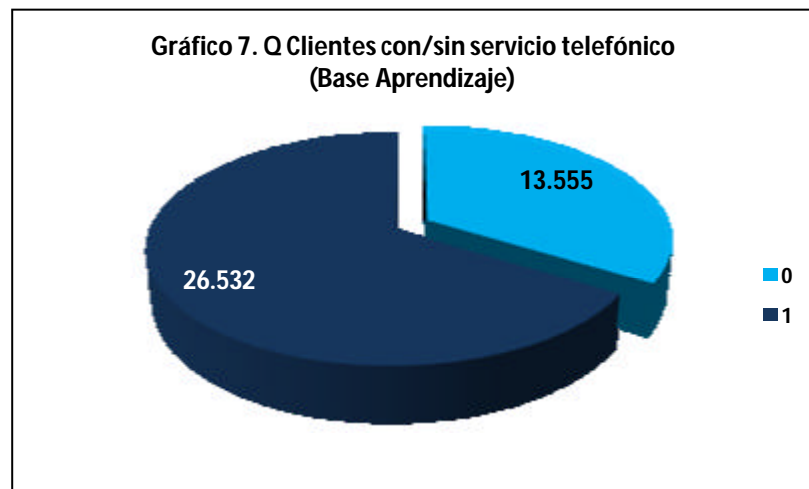


Fuente: Empresa

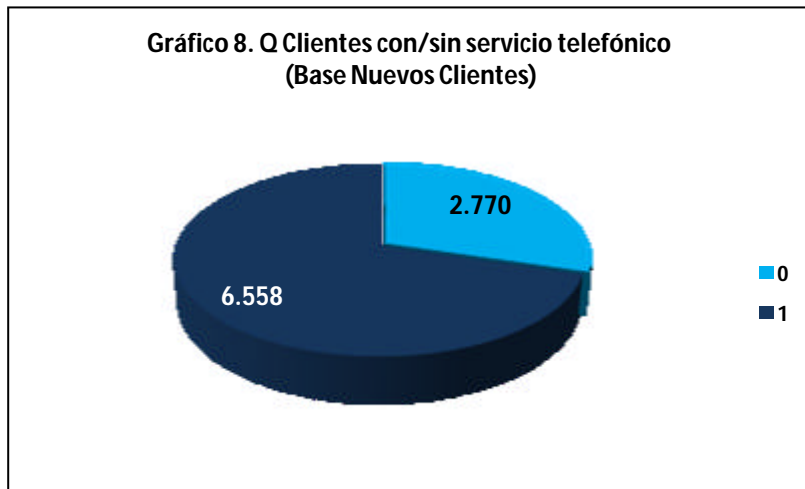
- c) **Servicio telefónico en plan base:** Variable cualitativa dicotómica. Esta variable indica si dentro del plan base contratado por el cliente, éste incluye o no telefonía.

La justificación entregada por la empresa para incluir esta variable en el modelo pasa por el hecho que un cliente al poseer telefonía, produce un efecto variable en el monto mensual a pagar, detectándose actualmente como una variable que el cliente justifica al momento de caer en mora. Un cliente puede excederse en su cuenta por llamadas LDN – larga distancia nacional -, LDI – larga distancia internacional-, celular, o bien, SLM – servicio local medido-.

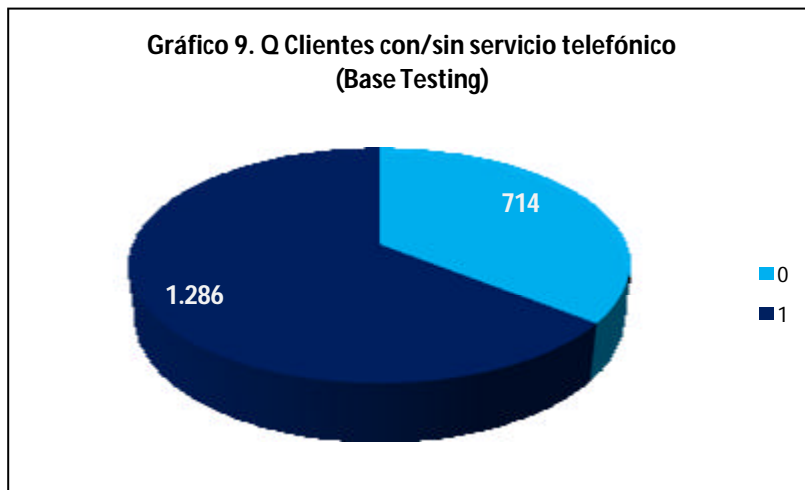
Se consideran dos modalidades: 1 si plan base posee telefonía y 0 si plan base no la posee.



Fuente: Empresa



Fuente: Empresa

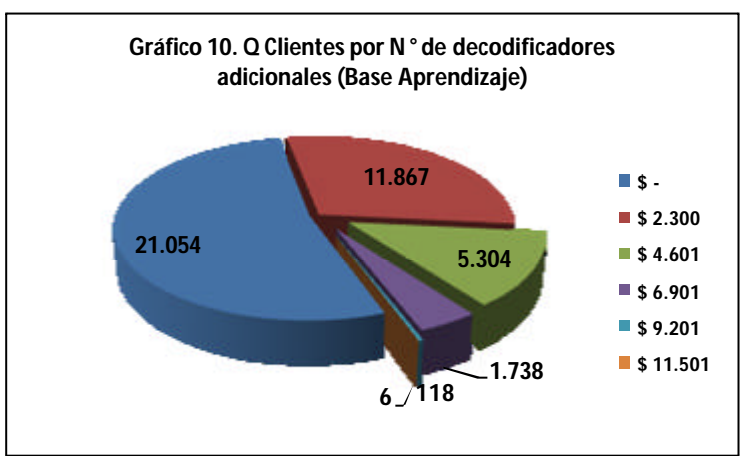


Fuente: Empresa

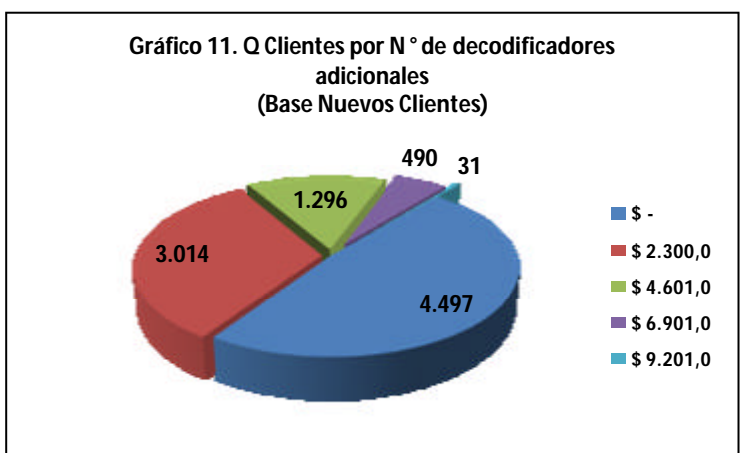
- d) **Decodificadores adicionales:** Variable cuantitativa discreta. Cada plan base que contiene televisión, incluye sin costo para el cliente un decodificador, no obstante, si se desea contar con el servicio de televisión en un segundo o tercer televisor en el hogar, se deben contratar decodificadores adicionales, los cuales tienen un costo unitario mensual por concepto de arriendo de \$2.300 (con IVA). Dentro de los servicios adicionales ofrecidos por la empresa, el más demandado son los decodificadores, por tanto, se ha determinado que esta variable debe considerarse en el modelo debido que aumenta el valor mensual que el

cliente debe cancelar por el servicio contratado y condiciona la morosidad.

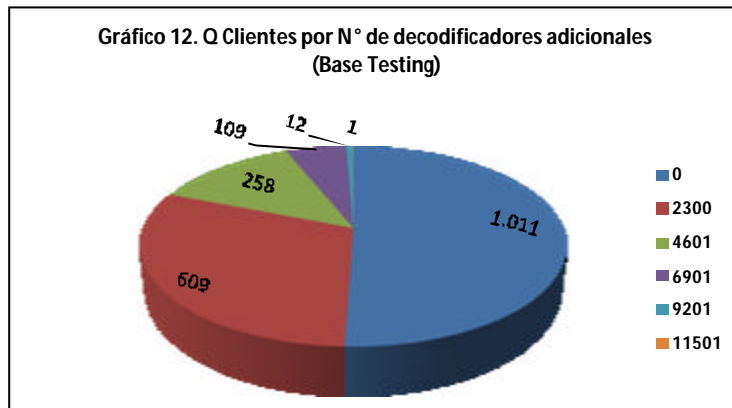
Esta variable considera entonces 6 modalidades, \$0 sin decodificador adicional, \$2.300 un decodificador adicional, \$4.601 dos decodificadores, \$6.901 3 decodificadores, \$9.201 4 decodificadores y \$11.501 5 decodificadores.



Fuente: Empresa.



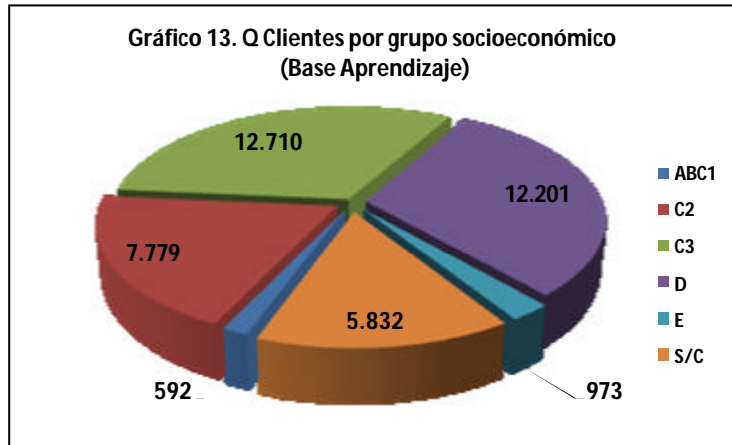
Fuente: Empresa.



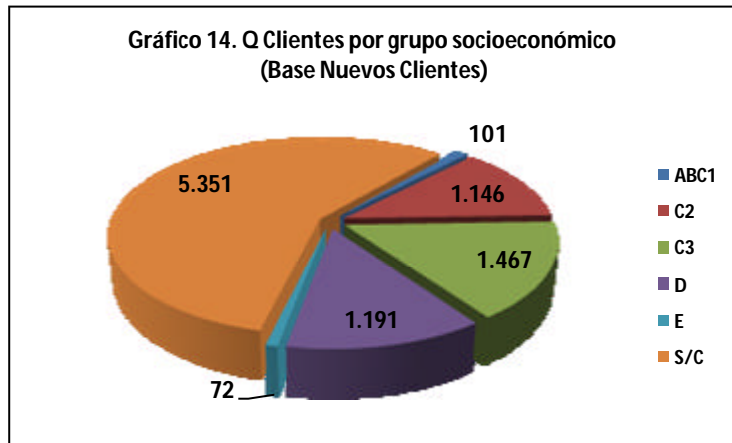
Fuente: Empresa.

- e) **GSE**: Variable cualitativa ordinal. Esta variable representa el grupo socio económico al cual pertenece el cliente, vale decir, si éste pertenece al segmento denominado como ABC1 ó C2 ó C3 ó D ó E. Dado a que el servicio es vía cable, la red está georeferenciada, permitiendo asociar información de este tipo. La empresa justifica que el grupo socioeconómico debe ser incluido porque entrega una idea de la capacidad de pago del cliente.

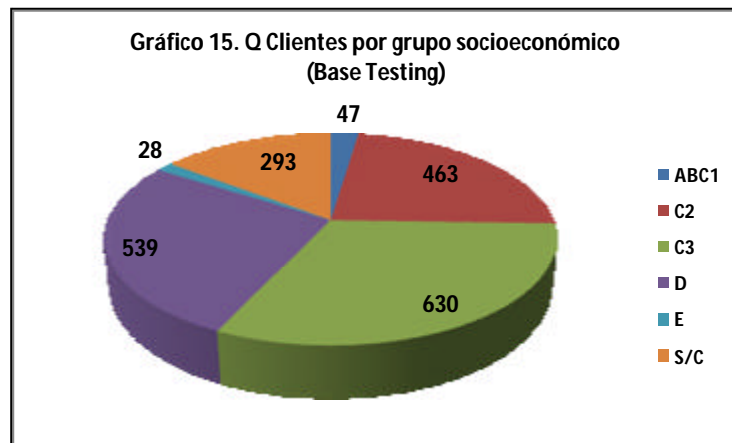
Las modalidades son seis, siendo estas ABC1, C2, C3, D, E, y una última, a la cual se ha codificado como S/C que identifica aquellas direcciones de clientes que no son factibles de asociar a un grupo socioeconómico.



Fuente: Empresa.



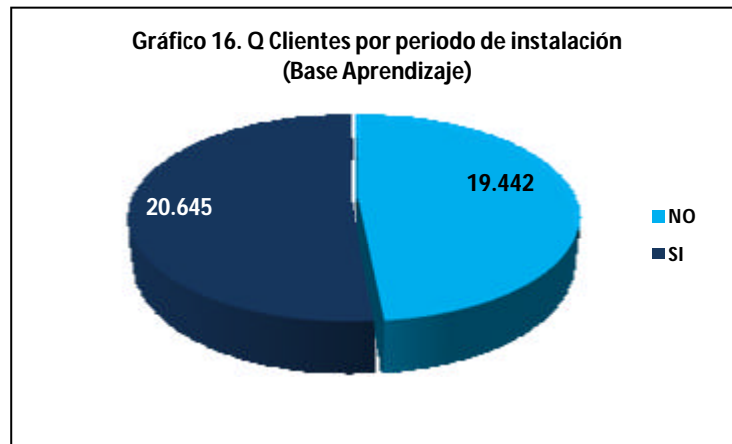
Fuente: Empresa.



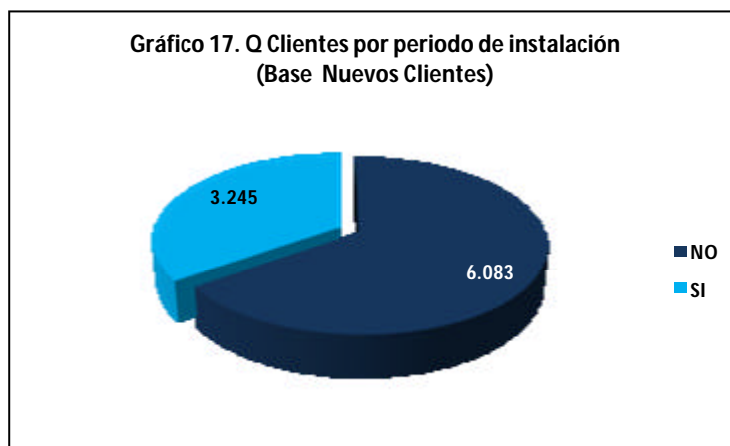
Fuente: Empresa.

f) **Instalación antes del día 15 del mes:** Variable cualitativa dicotómica. Debido al ciclo de facturación de la empresa, si una venta es realizada e instalada antes del día 15 (inclusive), el servicio será facturado y emitido en el mismo mes, siendo su vencimiento en el mes siguiente, día 4. Caso contrario, el servicio será emitido y facturado en el mes siguiente, siendo su vencimiento en el mes subsiguiente. La presencia de esta variable en el modelo se debe porque en el primer caso, el valor a cancelar por el cliente es proporcional, en cambio, en los casos de clientes con vencimiento en el mes subsiguiente a su instalación, el servicio será cobrado en su totalidad, es decir, el ciclo de facturación completo.

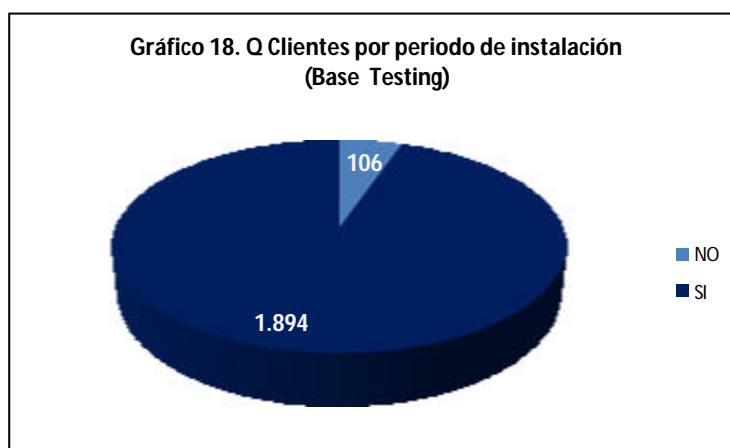
Esta variable tiene dos modalidades, Si - instalado antes del 15 inclusive -, No – servicio instalado con posterioridad al 15 del mes -.



Fuente: Empresa.

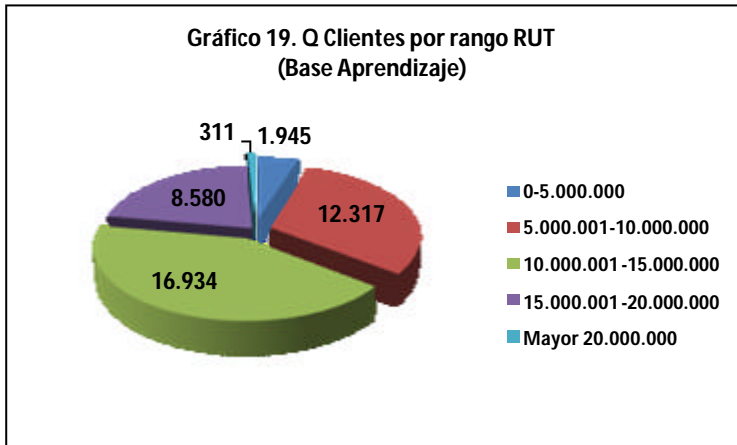


Fuente: Empresa.

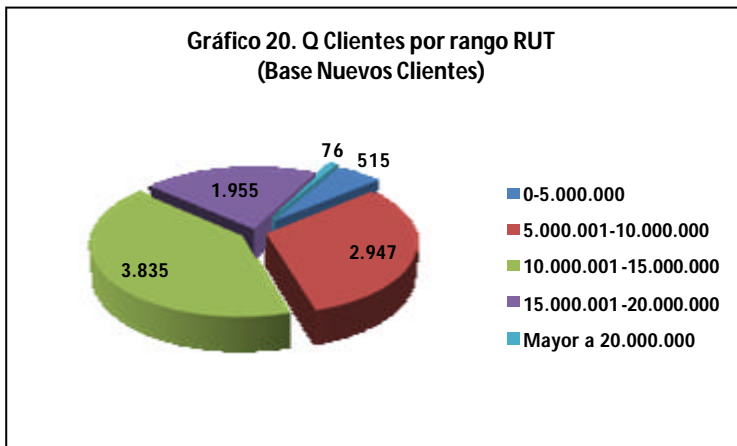


Fuente: Empresa.

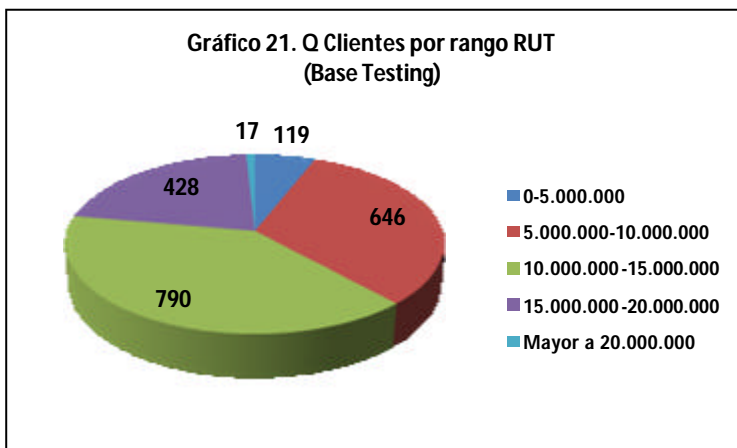
- g) **Rut:** Variable cuantitativa discreta. Al momento de hablar de morosidad, es sabido que la edad tiene bastante que decir en la responsabilidad de pago, dado que ésta no se posee como dato debido a que no es una variable de ingreso al sistema de la empresa al momento de contratar, se ha considerado tener una relación a través del RUT del cliente. La variable predictora RUT estará compuesta de 5 modalidades, rut menores a 5.000.000, entre 5.000.000 y 10.000.000, mayores a 10.000.000 y menores a 15.000.000, entre 15.000.000 y 20.000.000, por último, mayores a 20.000.000.



Fuente: Empresa.

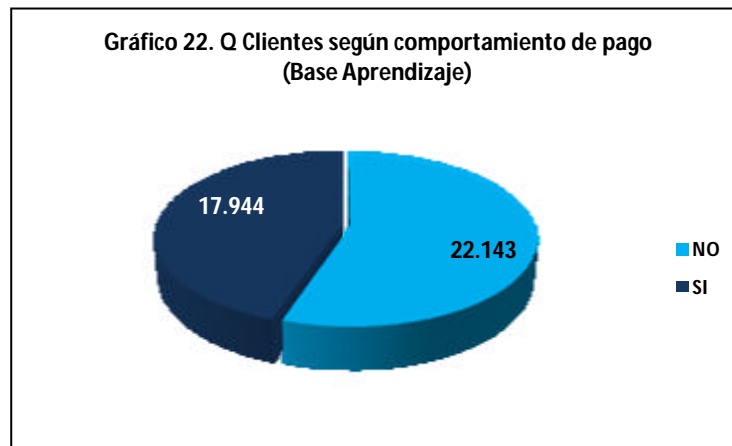


Fuente: Empresa.

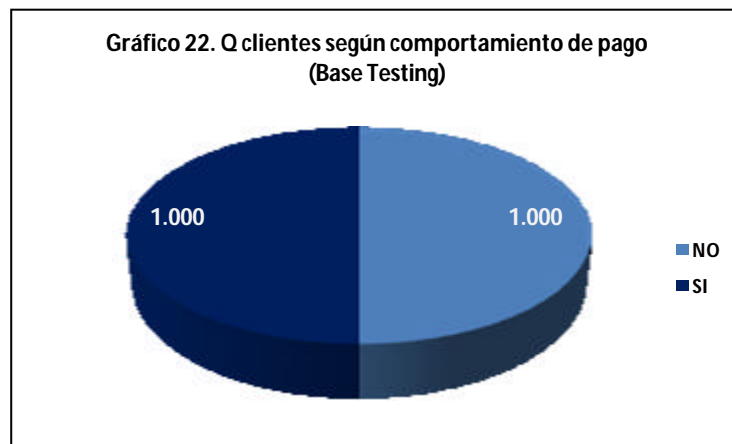


Fuente: Empresa.

h) **Paga al vencimiento:** Variable cualitativa dicotómica. Esta que corresponderá a la variable discriminante, y por tanto, sólo se posee información para los clientes que componen la base de aprendizaje y de testing. Es la variable discriminante dado que indica si un cliente pagó o no su primera factura al vencimiento.



Fuente: Empresa.



Fuente: Empresa.

3.3. Programación Modelo de Clasificación Bayesiana

En esta sección se explicará paso a paso el algoritmo de programación del modelo de clasificación bayesiana, las sentencias utilizadas, siendo esta sección la base y guía para los programadores de la empresa, de ahí la relevancia de indicar en detalle el algoritmo de programación.

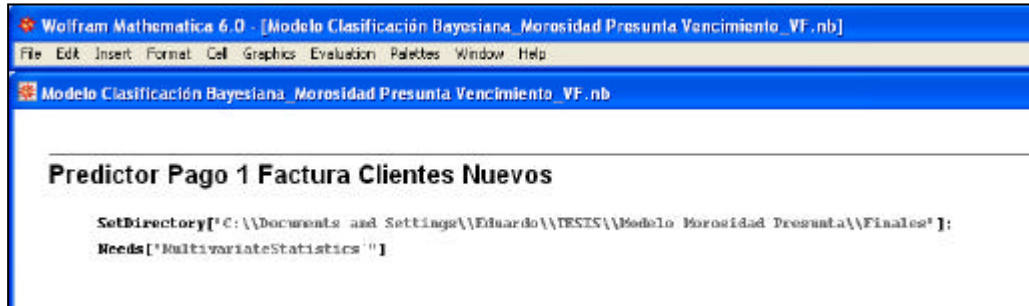
En el desarrollo de este capítulo se hablará de base de aprendizaje y de nuevos clientes, no obstante, resulta importante aclarar que en la fase de validación del modelo (testeo), la base de los nuevos clientes corresponderán a los que conforman la base de testing.

3.3.1. Definición del Directorio de Origen de las Bases de Datos

El paso inicial en la programación del modelo en *Wolfran Mathematica 6.0*, pasa por indicar el directorio en el cual se almacenan las bases de datos (aprendizaje y nuevos clientes), con ello, el programa sabrá hacia dónde dirigirse para extraer los datos y realizar los cálculos necesarios a través de las rutinas programadas.

Junto con lo anterior, en esta primera sección de la programación, se debe indicar a que paquete estadístico de la librería de *Wolfran Mathematica 6.0* estarán asociadas las funciones matemáticas a utilizar en el modelo. En el caso de clasificación Bayesiana, se utilizará *Multivariate Statistics Package*, el cual contiene estadística descriptiva para datos multivariados, distribuciones derivadas de la distribución normal multivariada y distribuciones discretas multivariadas.

Fig. 3 Sentencias de definición directorio de datos almacenados y definición de paquete estadístico.



Fuente: Elaboración propia.

??????????????????. Indica el nombre del directorio sobre el cual se encuentran almacenadas las bases de datos.

??????????????????. Indica el paquete estadístico a utilizar dadas las funciones que se programarán en el modelo.

3.3.2. Importación de Bases de Datos o Matrices

Se ha declarado el directorio sobre el cual se alojan las bases o matrices, ahora corresponde importarlas, depurarlas y asociarlas a unas variables, las cuales permitirá n identificarlas a lo largo de la programación.

Las sentencias para importar, depurar y asociar cada una de las matrices se describen a continuación:

?? ???????????????????. Función que importa el archivo a utilizar, se debe indicar el nombre del archivo y la extensión de éste, es decir, .xls en el caso del modelo dado que las bases están en formato Excel. La importación se realiza a modo de lista, es decir, cada fila de la matriz es un elemento de la lista, dónde el primer elemento de la lista (primera fila de la matriz) corresponde a los nombres de las variables (etiquetas ó campos), y los siguientes, a cada cliente y su respectivo vector característico.

Fig. 4 Sentencias de importación, depuración y renombramiento de bases de datos.

```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb

Carga Matriz Aprendizaje (Y) y Matriz Nuevos Clientes (Z)

Y = Flatten[Import["Base de Datos_BDD_Julio_Octubre_MATEMATICA.xls"], 1];
(*Flatten: Funcion que elimina las llaves interiores, Flatten[{a},{b},{c}];{d},1], {a,b},{c},d*)
Y = Drop[Y, {1}];
(*Drop: Funcion que elimina el primer elemento de la lista, ie,
el nombre de las variables*)

Z = Flatten[Import["Base de Datos_BDD_Pronostico Pago Vencimiento Diciembre 2009_MFC_MATEMATICA.xls"], 1];
Z = Drop[Z, {1}];
    
```

Fuente: Elaboración propia.

Para verificar que las sentencias descritas se han ejecutado correctamente para ambas matrices, es necesario realizar una validación, para ello, basta tomar el primer elemento de cada matriz y comprobar que la información sea exactamente igual a la del primer cliente del archivo Excel para cada matriz.

El primer elemento o cliente de la matriz de aprendizaje corresponde a un cliente **?? ?????????? ??????? ???? ???????????????????**, es decir, este vector característico del cliente indica que éste contrató vía terreno, el valor del plan base contratado es de \$25.501, posee telefonía en su plan, sin decodificador adicional, el segmento económico al cual pertenece es el D, su servicio fue instalado antes del día 15, su rut es el 10.383.361-2 y pagó su primera factura del servicio al vencimiento.

Fig. 5 Vector característico del primer cliente de la base de aprendizaje.

	A	B	C	D	E	F	G	H	I	J	K
1	Canal de Venta	valor plan	Serv tel	\$ Deco	SGE	INST <=15	RUT	PAGA 1 FACT			
2	INDIRECTA	25501	1	0 D	SI		103833612	SI			
3	INDIRECTA	15000	0	0 S/C	SI		109014222	SI			
4	INDIRECTA	15000	0	0 D	SI		139914120	SI			

Fuente: Elaboración propia.

El primer elemento de la matriz de nuevos clientes a clasificar, corresponde a un cliente que posee como vector característico a uno de la forma $?? ?????????? ?????? ?? ??^2 ?? ? ???????????$, es decir, la venta fue realizada en terreno, el valor del plan base contratado es de \$14.500, éste posee telefonía, sin decodificador adicional contratado, sin segmento económico asociado ó definido, fue instalado posterior a un día 15, su rut es el 4.623.583-5.

Fig. 5 Vector característico primer cliente de la base de nuevos clientes.

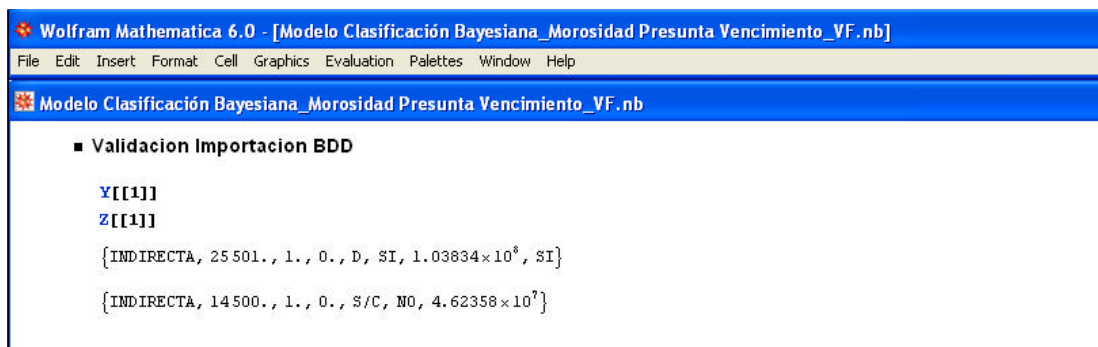
	A	B	C	D	E	F	G	H	I	J	K	L
1	Canal de Venta	valor plan	Serv tel	\$ Deco	SGE	INST <=15	RUT					
2	INDIRECTA	14500	1	0 S/C	NO		46235835					
3	REMOTA	14500	1	0 S/C	NO		35935697					
4	INDIRECTA	14500	1	0 S/C	NO		138788997					

Fuente: Elaboración propia.

Dado que la matriz de aprendizaje se ha renombrado como ? y la de nuevos clientes como ?, al realizar el comando $????? ???? ?$, Wolfram Mathematica

desplegará el vector característico del primer elemento de la lista para ambas matrices, es decir, la primera fila de cada matriz importada.

Fig. 6 Validación importación y depuración de bases.



Fuente: Elaboración propia.

Se valida que la importación del programa se ha realizado correctamente.

3.3.3. Codificación Bases de Aprendizaje y de Nuevos Clientes

Hasta el momento, las modalidades de cada una de las variables predictoras y discriminante están en algún formato ya sea del tipo texto, números, o números únicos como es el caso del rut – único para cada cliente - y ante la necesidad de realizar más adelante operaciones sobre valores numéricos enteros positivos, es necesario codificar el valor de las modalidades de cada variable, para ello, se realizan las siguientes fases acompañadas de sus respectivas sentencias.

3.3.3.1. Generación del procedimiento de codificación

Sea ???????? ??,?, el nombre del procedimiento de codificación de modalidades de cada variable.

La sentencia para definir un procedimiento (procedure) en *Wolfram Mathematica* corresponde a `Module[{local variables}, body]`, en el modelo, las variables locales corresponderán a `{local variables}` las cuales se describen a continuación, y el proceso será la codificación propiamente tal de cada variable.

Sean las siguientes variables locales,

`Y`: Variable que tomará el valor `Z`, siendo éste el argumento de la función `Codifica[Z_]`, en el modelo, `Codifica[Z_]` puede ser `Module[{n, k}, ...]`, esto dependiendo de la matriz que en ese momento se desea codificar.

`n`: Variable local que estará asociada al largo de la variable local `Y`, es decir, tendrá el valor 40.087 si `Z` ó 9.328 si `Z`, correspondiendo estos valores a la cantidad de cliente que contiene cada matriz.

`k`: Variable local que almacenará el largo del primer elemento de la variable local `Y`, dependiendo de la matriz que se esté codificando, será equivalente al número de variables, por ejemplo, la matriz de aprendizaje contiene 8 variables, 7 predictoras más la variable discriminante paga o no paga. En el caso de la matriz de nuevos clientes tomará el valor 7 dado que la variable discriminante es la que se requiere determinar para estos clientes.

Fig. 7 Sentencia creación procedimiento de codificación y definición de sus variables locales.

```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb
Codificación Matrices Aprendizaje y Nuevos Clientes

Codifica[Z_] := Module[{Y, n, k},
  Y = Z;
  n = Length[Y];
  k = Length[Y[[1]]];

```

Fuente: Elaboración propia.

3.3.3.2. Definición de códigos para variables predictoras y discriminante.

Para cada una de las variables predictoras como también para la variable discriminante, se indica en esta sección la codificación definida y las sentencias encargadas de realizarla.

Codificación modalidades de la variable canal de venta

Para facilitar la comprensión del proceso de codificación de cada variable, es necesario reforzar nuevamente que cada matriz es representada por filas (elementos, es decir, clientes) y por columnas (variables), no obstante, Wolfram Mathematica, la importación y trabajo sobre una matriz lo realiza a través de un formato del tipo lista, es decir, trabaja con las filas, a modo de ejemplo explicativo, supongamos que la matriz de aprendizaje está formada por dos clientes y cada cliente posee una modalidad para cada una de las 8 variables que lo caracteriza, es decir, se tiene una matriz 2×8 . Al desplegarla en el programa, éste la mostraría de la siguiente manera,

$\{\{1, 2, 3, 4, 5, 6, 7, 8\}, \{9, 10, 11, 12, 13, 14, 15, 16\}\}$

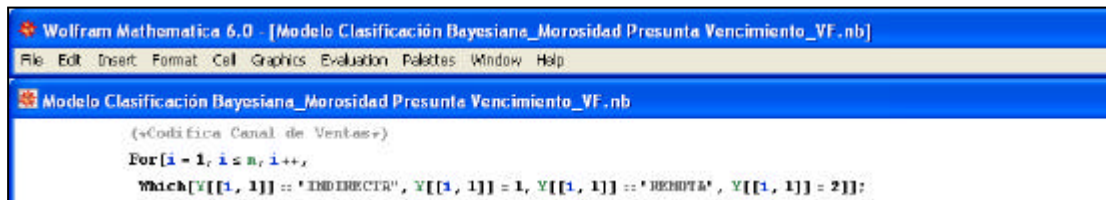
Es decir, una lista de dos elementos, donde cada elemento posee un vector característico de 8 coordenadas, esta es la forma o estructura con la cual se trabaja en el programa.

Aclarado lo anterior, para codificar la variable predictora, canal de ventas, se deberá recorrer cada elemento de la lista, para ello, se programa la sentencia `Do[...]`, que recorrerá desde el primer al último elemento de la matriz que se está codificando, ya sea esta la de aprendizaje o la de nuevos clientes. En cada elemento, se fijará en la primera variable del vector

característico del elemento, es decir, canal de ventas, y preguntará si su modalidad es igual a “INDIRECTO” ó “REMOTA”, asignándole el valor 1 ó 2 respectivamente según sea el caso.

La sentencia condicional a utilizar para determinar qué valor asignar como código corresponderá a `Which[test1 == "INDIRECTO", valor1, test2 == "REMOTA", valor2, ...]`, básicamente esta función valida si el test1 es verdadero, entonces le asigna el valor1, caso contrario, aplica el test2 y asigna el valor2, y así sucesivamente.

Fig. 8 Sentencias de codificación variable de aprendizaje Canal de Ventas.



```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb
(*Codifica Canal de Ventas*)
For[i = 1, i <= n, i++,
  Which[Y[[i, 1]] == "INDIRECTO", Y[[i, 1]] = 1, Y[[i, 1]] == "REMOTA", Y[[i, 1]] = 2];
  
```

Fuente: Elaboración propia.

Codificación de modalidades de la variable plan base

La variable plan base, corresponde a la segunda coordenada del vector característico de cada elemento de la matriz, ya sea ésta la de aprendizaje o la de nuevos clientes. Esta variable posee 11 modalidades, las cuales serán codificadas desde el plan de menor valor al de mayor valor con los números del 1 al 11 respectivamente.

Fig. 9 Sentencias de codificación modalidades variable de aprendizaje Plan base.

```
Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb
(*Codifica Canal de Ventas*)
For[i = 1, i ≤ n, i++,
  Which[Y[[i, 1]] == "INDIRECTA", Y[[i, 1]] = 1, Y[[i, 1]] == "REMOTA", Y[[i, 1]] = 2]];
(*Codifica Valor del Plan Base*)
For[i = 1, i ≤ n, i++, Y[[i, 2]] = IntegerPart[Y[[i, 2]]]];
For[i = 1, i ≤ n, i++,
  Which[Y[[i, 2]] == 14500, Y[[i, 2]] = 1,
  Y[[i, 2]] == 15000, Y[[i, 2]] = 2,
  Y[[i, 2]] == 20500, Y[[i, 2]] = 3,
  Y[[i, 2]] == 21500, Y[[i, 2]] = 4,
  Y[[i, 2]] == 24500, Y[[i, 2]] = 5,
  Y[[i, 2]] == 25501, Y[[i, 2]] = 6,
  Y[[i, 2]] == 27000, Y[[i, 2]] = 7,
  Y[[i, 2]] == 29000, Y[[i, 2]] = 8,
  Y[[i, 2]] == 34500, Y[[i, 2]] = 9,
  Y[[i, 2]] == 37500, Y[[i, 2]] = 10,
  Y[[i, 2]] == 42000, Y[[i, 2]] = 11]];

```

Fuente: Elaboración propia.

Codificación modalidades de la variable predictora telefonía.

Un plan base puede o no estar compuesto por el servicio de telefonía, lo cual se refleja en las bases origen en el campo Serv tel como 1 ó 0 respectivamente. Dado que se debe trabajar con números enteros positivos para efectos de codificación en el programa, se codificará como 1 aquel cliente que posee telefonía y como 2 aquel cliente que no posee este servicio en su plan base.

Dentro de cada elemento de la lista, esta variable ocupa la coordenada número tres del vector característico, dado lo anterior, se programan las sentencias tanto para recorrer cada elemento de lista de cada una de las matrices, aprendizaje y nuevos clientes, según variable de entrada ?, como también, para preguntar, fijándose en la coordenada 3 del elemento (vector), si se posee o no telefonía en el plan base para ser codificado como 1 ó 2 respectivamente.

Fig. 10 Sentencias de codificación variable de aprendizaje Telefonía en plan base.

```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb
(*Codifica si posee o no telefonía*)
For[i = 1, i ≤ n, i++, Y[[i, 3]] = IntegerPart[Y[[i, 3]]]:
For[i = 1, i ≤ n, i++,
Which[Y[[i, 3]] == 0, Y[[i, 3]] = 2, Y[[i, 3]] == 1, Y[[i, 3]] = 1]]:
    
```

Fuente: Elaboración propia.

Codificación modalidades variable predictor de decodificadores adicionales.

Un cliente que ha contratado televisión dentro de su plan base, puede haber o no agregado decodificadores adicionales, los cuales tienen un valor unitario adicional de \$2.300 (IVA incluido). Dentro de cada elemento o cliente, la variable decodificador adicional corresponde a la cuarta coordenada de su vector característico. Para efectos de codificación, se ha determinado la siguiente asociación:

```

??? ?????????????? ????? ?
?? ?????????????? ???Q???? ?
??? ?????????????? ???Q???? ?
???? ?????????????? ???Q???? ?
?????? ? ? ?? ???????????????? ??Q???? ?
    
```

Fig. 11 Sentencias de codificación variable de aprendizaje Decodificadores adicionales.

```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb
(*Codifica valor de decodificadores adicionales*)
For[i = 1, i ≤ n, i++, Y[[i, 4]] = IntegerPart[Y[[i, 4]]]];
For[i = 1, i ≤ n, i++,
  Which[Y[[i, 4]] == 0, Y[[i, 4]] = 1,
    Y[[i, 4]] == 2300, Y[[i, 4]] = 2,
    Y[[i, 4]] == 4601, Y[[i, 4]] = 3,
    Y[[i, 4]] == 6901, Y[[i, 4]] = 4,
    Y[[i, 4]] ≥ 9201, Y[[i, 4]] = 5]];

```

Fuente: Elaboración propia.

Codificación modalidades de la variable predictora segmento económico.

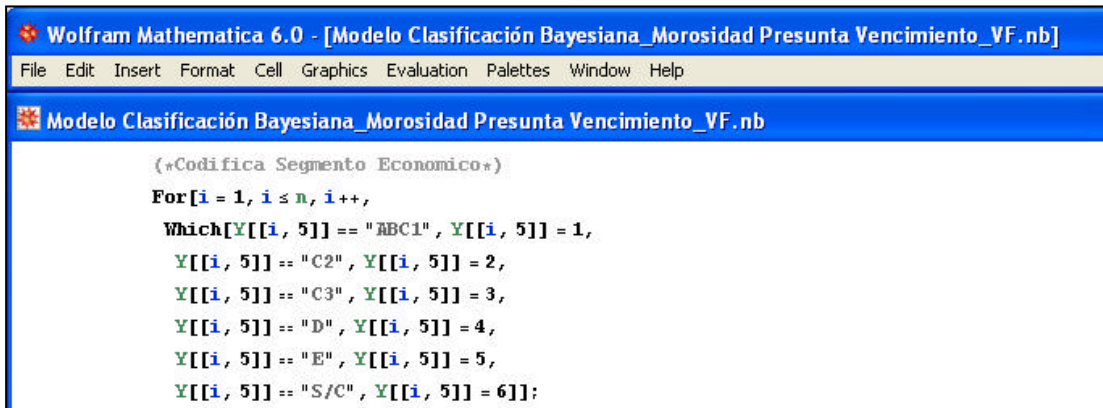
La variable predictora segmento económico, corresponde para cada elemento de la lista (cliente), a la coordenada número cinco del vector característico de un cliente. Esta variable posee seis modalidades, y se ha optado por la siguiente codificación:

```

???? ???? ????? ?
???? ???? ?? ? ?
???? ???? ?? ? ?
???? ???? ? ? ?
???? ???? ? ? ?
??? ???? ???? ? ?

```

Fig. 12 Sentencias de codificación variable de aprendizaje Segmento económico.



```
Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb

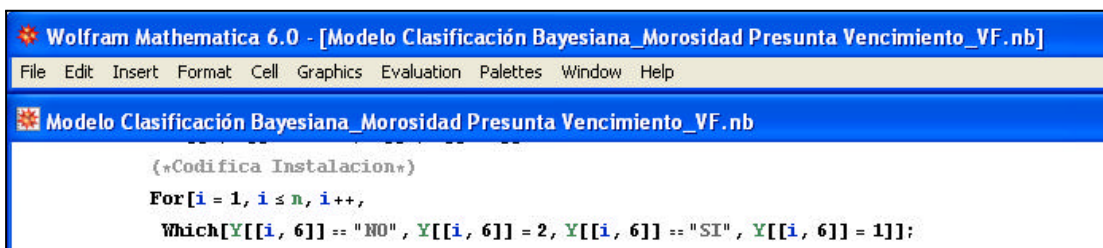
(*Codifica Segmento Economico*)
For[i = 1, i ≤ n, i++,
  Which[Y[[i, 5]] == "ABC1", Y[[i, 5]] = 1,
    Y[[i, 5]] == "C2", Y[[i, 5]] = 2,
    Y[[i, 5]] == "C3", Y[[i, 5]] = 3,
    Y[[i, 5]] == "D", Y[[i, 5]] = 4,
    Y[[i, 5]] == "E", Y[[i, 5]] = 5,
    Y[[i, 5]] == "S/C", Y[[i, 5]] = 6];
```

Fuente: Elaboración propia.

Codificación modalidades variable predictora instalación del servicio.

La fecha de instalación, en el sentido si el servicio fue instalado con anterioridad al día 15 del mes (inclusive) o no, como se indicaba anteriormente, impacta en cuarto a si el servicio será cobrado proporcionalmente o no en su primera factura. Esta variable, para cada elemento, se ubica en la coordenada número seis de su vector característico. Se codificará como 1 si fue instalado antes del día 15 (inclusive), y como 2 aquel servicio instalado con posterioridad.

Fig. 13 Sentencias de codificación variable de aprendizaje Instalación antes del día 15 del mes.



```
Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb

(*Codifica Instalacion*)
For[i = 1, i ≤ n, i++,
  Which[Y[[i, 6]] == "NO", Y[[i, 6]] = 2, Y[[i, 6]] == "SI", Y[[i, 6]] = 1];
```

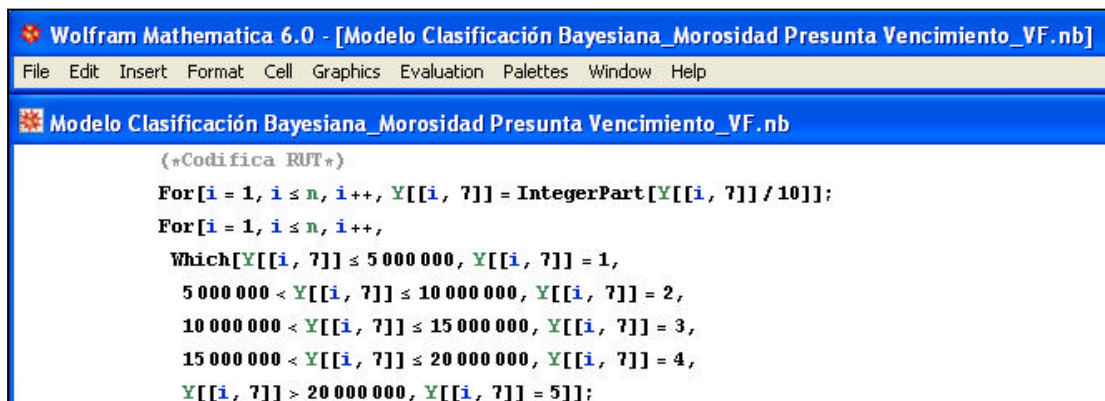
Fuente: Elaboración propia.

Codificación modalidades variable predictora RUT clientes.

Esta variable, que ocupa la coordenada número siete del vector característico de cada elemento de la lista, tanto en la matriz de aprendizaje como la de nuevos clientes, será codificada por rangos, siendo estos:

```
???? ?Q??Q?? ? ?  
?Q??Q?? ? ??? ? ?Q??Q?? ? ?  
??Q??Q?? ? ??? ? ?Q??Q?? ? ?  
??Q??Q?? ? ??? ? ?Q??Q?? ? ?  
???? ? ?Q??Q?? ? ?
```

Fig. 14 Sentencias de codificación variable de aprendizaje RUT.



```
Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]  
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help  
Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb  
(*Codifica RUT*)  
For[i = 1, i ≤ n, i++, Y[[i, 7]] = IntegerPart[Y[[i, 7]] / 10];  
For[i = 1, i ≤ n, i++,  
  Which[Y[[i, 7]] ≤ 5 000 000, Y[[i, 7]] = 1,  
        5 000 000 < Y[[i, 7]] ≤ 10 000 000, Y[[i, 7]] = 2,  
        10 000 000 < Y[[i, 7]] ≤ 15 000 000, Y[[i, 7]] = 3,  
        15 000 000 < Y[[i, 7]] ≤ 20 000 000, Y[[i, 7]] = 4,  
        Y[[i, 7]] > 20 000 000, Y[[i, 7]] = 5];
```

Fuente: Elaboración propia.

Codificación modalidades de la variable discriminante.

Por último queda por codificar la variable paga o no al vencimiento, dado que sólo está presente en la matriz de aprendizaje del modelo, ocupando la coordenada número ocho del vector característico de cada elemento, antes de su codificación, es necesario en el procedimiento que se realice la pregunta si la variable largo del elemento de la matriz que se está codificando en ese momento es igual a 8, si es así, el procedimiento sabe que está realizando acciones sobre la matriz de aprendizaje y por tanto codifica la variable paga o

Fig. 16 Sentencia de validación del procedimiento de codificación.

```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb

Z1 = Codifica[Y]; (*Matriz Aprendizaje*)
Z2 = Codifica[Z]; (*Matriz Nuevos Clientes*)

Z1[[1]]
Z2[[1]]

{1, 6, 1, 1, 4, 1, 3, 1}
{1, 1, 1, 1, 6, 2, 1}

{1, 6, 1, 1, 4, 1, 3, 1}
{1, 6, 1, 1, 4, 1, 3, 1}
    
```

Fuente: Elaboración propia.

De igual forma es validado para el caso de nuevos clientes.

3.3.4. Cálculo de la Probabilidad Bayesiana

En esta sección, se programa la parte principal del modelo, esto es el cálculo de la probabilidad que un nuevo cliente X , definido por su vector característico, no pague su primera boleta al vencimiento de ésta, y por tanto, caiga en morosidad de su servicio.

Sea $X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ el nuevo cliente caracterizado por las modalidades $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ asumidas en cada una de las siete variables y x_8 la variable paga, con sus dos modalidades si paga, no paga.

Interesa calcular $P(Y=1|X)$ y $P(Y=2|X)$, esto es la probabilidad que el cliente X pague y la probabilidad que el cliente X no pague. Como $P(Y=1|X) + P(Y=2|X) = 1$ es suficiente calcular $P(Y=1|X)$. Usando las propiedades de las probabilidades condicionales se tiene que:

$$P(g|c) = \frac{P(c|g) \cdot P(g)}{P(c)}$$

Donde $P(c|g)$ es la probabilidad de que elegido un cliente c este tenga las modalidades (g_1, g_2, \dots, g_n) . Además se sabe que $P(c) = \sum_{g=1}^n P(c|g) \cdot P(g)$ donde $P(g)$ es la probabilidad de que al elegir un cliente, éste tenga la modalidad g en la variable discriminante, luego

$$P(g|c) = \frac{P(c|g) \cdot P(g)}{\sum_{g=1}^n P(c|g) \cdot P(g)}$$

Por lo tanto, es necesario calcular las probabilidades condicionales $P(c|g)$, $P(g|c)$ y las probabilidades $P(c)$, $P(g)$. Usando la base de aprendizaje, y considerando independencia condicional en las variables [ver 3.1.5], se tiene:

$$P(c|g) = P(c_1|g) \cdot P(c_2|g) \cdot \dots \cdot P(c_n|g)$$

Por lo tanto

$$P(c) = \sum_{g=1}^n P(c_1|g) \cdot P(c_2|g) \cdot \dots \cdot P(c_n|g) \cdot P(g)$$

$$P(g|c) = \frac{P(c_1|g) \cdot P(c_2|g) \cdot \dots \cdot P(c_n|g) \cdot P(g)}{\sum_{g=1}^n P(c_1|g) \cdot P(c_2|g) \cdot \dots \cdot P(c_n|g) \cdot P(g)}$$

Explicado lo anterior, se describen ahora cada una de las etapas de la programación para el cálculo de la probabilidad bayesiana de cada nuevo cliente. [10]

Definición de variables

Previo a la realización de las sentencias encargadas de realizar el cálculo de la probabilidad Bayesiana, se debe primeramente transponer la matriz de aprendizaje, de tal forma que, las filas ahora representen las variables (7

predictoras, 1 discriminante) y las columnas representen a cada elemento ó cliente de la matriz, dónde entonces $matrix_{ij}$ hace referencia al cliente i y al valor de la modalidad de la variable j de su vector característico. La necesidad de transponer se debe a que se trabajará sobre las variables, y como se ha indicado, el programa realiza acciones sobre las filas de una matriz (listas).

Realizada la transpuesta de la matriz de aprendizaje codificada ($matrix$), y renombrada ahora como $matrix^T$, se hace necesario ahora definir las variables que serán llamadas en las sentencias de cálculo siguientes para el cálculo de la probabilidad, por tanto sea:

$matrix^T_{ij}$: Vector que contiene como elementos el número de modalidades asociada a cada una de las variables predictoras, en el modelo $matrix^T_{ij}$.

$matrix^T_{ij}$: Variable que representa el número de la coordenada que ocupa la variable discriminante en el vector característico de cada cliente que compone la matriz de aprendizaje, en el modelo $matrix^T_{ij}$.

$matrix^T_{ij}$: Variable que contiene el número de modalidades que compone la variable discriminante, en el modelo $matrix^T_{ij}$.

$matrix^T_{ij}$: Vector que contiene como coordenadas, el valor de cada una de las modalidades de la variable discriminante, es decir, en el modelo $matrix^T_{ij}$.

$matrix^T_{ij}$: Variable que toma como valor el número de variables predictoras, por tanto $matrix^T_{ij}$.

Fig. 17 Sentencia para transponer una matriz y definición de variables de cálculo.

The screenshot shows the Wolfram Mathematica 6.0 interface with the following code in a cell:

```

S = Transpose[Z1];

Mo = {2, 11, 2, 5, 6, 2, 5}; (*Vector modalidad de cada variable, no incluye var discriminante*)
nc = 8; (*numero de la variable discriminante*)
k = 2; (*numero de modalidades de la variable discriminante*)
c = {1, 2}; (*vector modalidad de la variable discriminante*)
r = Length[Mo];
(*contiene el número de variables*)
    
```

Fuente: Elaboración propia.

Casos de pagadores y no pagadores

Sea \mathbf{v} , un vector compuesto de dos coordenadas, la primera representa el valor del número de pagadores, y la segunda, contiene el número de no pagadores del total de clientes de la matriz de aprendizaje.

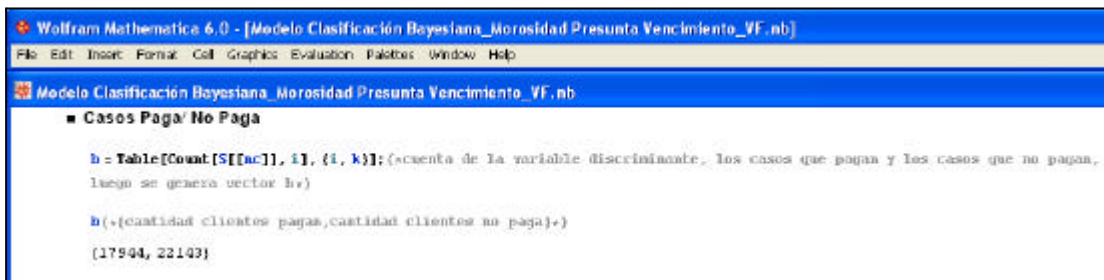
Con estos valores ya identificados y almacenados, será de manera fácil calcular la probabilidad de pagadores y no pagadores que se requiere, utilizando para ello la lógica de casos favorables sobre casos totales. En esta sección, por tanto, el modelo genera la información base para calcular las probabilidades $P(\text{pagador})$ y $P(\text{no pagador})$.

Para construir este vector \mathbf{v} , se utilizan dos sentencias en Mathematica, `Count` y `Length`, a continuación a través de un ejemplo se da a conocer la estructura y definición de cada una de ellas.

`Count` Función que cuenta el número de veces que el argumento `elemento` aparece como elemento en la matriz o lista llamada como `matriz`.

????? ?? ???? ?? Crea una tabla de la forma ??????????, es decir, cada valor del 1 al 6 lo evalúa en ? y almacena su resultado en una tabla.

Fig. 18 Sentencia para cálculo de pagadores y no pagadores de la matriz de aprendizaje.



Fuente: Elaboración propia.

Matriz de coincidencias entre modalidades de cada variable predictora y modalidades de la variable discriminante.

Un nuevo cliente, tiene un vector característico que lo identifica, dónde cada una de las variables predictoras de este vector toma un valor (modalidad), a modo de ejemplo, sea \vec{x} el vector característico de un cliente, el cual queda definido de la siguiente manera, $\vec{x} = (x_1, x_2, \dots, x_n)$, dónde la modalidad de su primera variable predictora llamada como canal de ventas, corresponde a indirecta –recordar que posee dos modalidades, indirecta y remota, codificadas como 1 y 2 respectivamente-; para su segunda variable predictora llamada como plan base, su modalidad toma el valor 25.501, y así sucesivamente.

Ahora bien, para calcular las probabilidades condicionales independientes entre cada modalidad, dado que pagó, como también, dado que no pagó, basándose para ello de la matriz de aprendizaje, se debe primeramente generar una función, la cual se denominará como $f(x)$, que tendrá como objetivo primeramente armar pares entre el valor de cada modalidad de cada variable

predictora de cada cliente, con cada valor de la modalidad de su variable discriminante, para luego, dado un par de la forma:

Contar cuantas coincidencias de este par existen en la matriz de aprendizaje.

La sentencia que realiza lo anteriormente descrito se programa a través de lo siguiente:

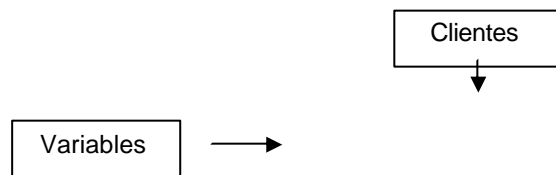
`funcion(?, ?, ?)`

La función `funcion`, requiere como entrada, un valor para cada una de sus tres variables, la primera variable, `?`, representa a una variable predictora, por ejemplo, si `?`, se estará haciendo referencia a la primera variable predictora del vector característico de un cliente, es decir, al canal de ventas; la variable `?` toma un valor comprendido entre 1 y el número total de modalidades de la variable predictora `?`, si `?`, se estará considerando la variable predictora canal de ventas y su modalidad denominada como indirecta, recordar que se ha codificado como 1 al canal de ventas indirecto, y como 2 al canal de ventas remoto. Por último, se tiene la variable denominada como `??` que puede tomar el valor 1 ó 2, valores que corresponden a la codificación realizada para las modalidades de la variable discriminante, pagador y no pagador respectivamente.

Dentro de `funcion`, la sentencia `matriz = matriz[?][?]`, siendo `?` (lugar que ocupa la variable discriminante en el vector característico de un cliente, ó en otras palabras, corresponde a la fila 8 de la matriz `matriz` - matriz de aprendizaje codificada transpuesta-, genera una matriz de $2 \times n$, donde la primera fila corresponde a la variable predictora, y la segunda a la discriminante, representando las columnas a los `n` clientes de la matriz de aprendizaje.

Para mayor comprensión de lo anteriormente descrito, supongamos que x_1 y x_2 son dos variables predictoras, cada una compuesta de dos modalidades, y sea y la variable discriminante también compuesta de dos modalidades. Supongamos ahora que la matriz de aprendizaje está compuesta por 3 clientes, y sus vectores característicos son de la forma:

Quedando entonces una matriz S de la siguiente forma:



Entonces, en el ejemplo, $S = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ y si $y = 1$, al ejecutar la sentencia `predict(S, y)`, se tendría la siguiente forma matricial,

Es decir se estaría asociando la modalidad de la primera variable predictora para cada uno de los clientes con el valor de la modalidad de sus respectivas variables discriminantes.

Considerando por un lado que se busca determinar el número de coincidencias entre una modalidad tanto con pagador como con no pagador en la matriz de aprendizaje para efectos del cálculo de las probabilidades

La construcción de esta matriz, se realiza mediante la programación de las siguientes sentencias:

Coin = Table[Aux[i, j, s], {s, 2}, {i, r}, {j, M}]]

Dónde i toma los valores desde 1 a r , donde r es el número total de variables predictoras; j toma desde el valor 1 a M , siendo M el número de modalidades que componen a la variable predictora j , y por último, s que tomará el valor de cada una de las modalidades de la variable discriminante codificada (1: pagador, 2: no pagador). La lógica del incremento del valor que toman las variables i será primeramente $i+1$, luego $j+1$ y finalmente $s+1$, una vez que s cumpla con su rango, i tomará su segundo valor, j tomará los valores desde 1 hasta el número de modalidades de la nueva variable predictora asociada al nuevo valor de i , y así sucesivamente hasta que $i=r$, última variable predictora, permitiendo entonces ahora a s tomar el valor 2 (no pagador), repitiéndose el ciclo descrito nuevamente.

Fig. 19 Sentencias de construcción matriz de coincidencias para cada variable de aprendizaje, entre sus modalidades con modalidades de variable discriminante (pagadores y no pagadores).

```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb

■ Cálculo Probabilidades Condicionales del vector característico de un cliente

Aux[w_, wj_, cl_] := Count[Transpose[{S[[w]], S[[w]]}], {wj, cl}];
(*Función que: Genera matrix de 2xn, entre una variable y la discriminante, la T, genera matrix nx2,
pares ordenados, luego se cuentan las coincidencias*)
Coin = Table[Aux[i, j, s], {s, 2}, {i, r}, {j, M}];
(*Genera Matrix de coincidencias de cada variable de su respectiva modalidad con pagadores y luego con no
pagadores*)

MatrixForm[Coin]

[[{14344, 3600} {226, 2795, 1277, 2478, 120, 2855, 1418, 1442, 741, 2746, 1845} {11568, 6376} {9098, 5544, 2490, 746, 66}
{18729, 3414} {323, 3200, 1424, 3156, 131, 3581, 1761, 1615, 809, 3504, 2639} {14964, 7179} {11956, 6323, 2814, 992, 58}]]

```

Fuente: Elaboración propia.

En la figura 19, al ejecutar el comando `print(X_train)`, el programa muestra la forma matricial que tiene la matriz `X`, la cual está formada por dos filas y 7 columnas. A modo de ejemplo, la fila 1 columna 1, representa cuantas coincidencias hay entre `X_train[0,0]` y `X_train[1,0]`, vale decir, de la matriz de aprendizaje 14.344 clientes vendidos a través del canal de ventas indirecta pagaron, y 3.600 clientes vendidos a través del canal de ventas remoto pagaron. Si tomamos la fila 2 columna 1, tendremos el mismo escenario, no obstante, con lo que no pagaron, siempre considerando al vencimiento.

Hasta el momento se cuenta con el vector `Y`, dónde `Y[0]` y `Y[1]`, el cual almacena la cantidad de clientes que pagan y no pagan respectivamente de la matriz de aprendizaje, como también, se cuenta con la matriz `X` que almacena la cantidad de veces que coincide cada valor de cada una de las modalidades de las variables predictoras con cada una de las dos modalidades de la variable discriminante en la matriz de aprendizaje.

Ahora solo resta programar y utilizar las sentencias adecuadas que permitan calcular la probabilidad bayesiana, es decir, calcular:

$$P(Y=0|X) = \frac{P(Y=0) \prod_{i=1}^7 P(X_i|Y=0)}{P(Y=0) \prod_{i=1}^7 P(X_i|Y=0) + P(Y=1) \prod_{i=1}^7 P(X_i|Y=1)}$$

Para ello, primeramente, se programaran las sentencias necesarias para calcular `P(Y=0|X)`, considerando independencia condicional de cada variable predictor, la probabilidad `P(Y=0|X)` será entonces la productoria de las probabilidades condicionales entre el valor de la modalidad de cada variable predictor dado que se pagó (la misma lógica se aplica para calcular `P(Y=1|X)`). Se busca entonces programar `P(Y=0|X)` y `P(Y=1|X)` g `P(Y=0|X)`. Se definirá para ello la función `bayes`, la cual se encargará de realizar la productoria de las probabilidades condicionales independientes, ésta tendrá

como variables de entrada a las variables x_i , siendo x_i el valor de una de las dos modalidades de la variable discriminante (1: paga, 2: no paga), y x el vector característico de un nuevo cliente, el cual posee un valor definido para cada modalidad para cada una de las 7 variables predictoras consideradas en el modelo. Entonces, la función $P(x)$ quedará definida como:

$$P(x) = \frac{P(x_1) \cdot P(x_2) \cdot \dots \cdot P(x_7)}{P(x_1) + P(x_2) + \dots + P(x_7)}$$

Ahora, se definirá la función $P(x)$, la que estará a cargo de calcular la probabilidad bayesiana de que un nuevo cliente, definido por su vector característico x pague su primera factura al vencimiento (el modelo también podrá calcular el caso dado que no paga). La función $P(x)$ en Wolfram Mathematica queda definida como:

$$P(x) = \frac{P(x_1) \cdot P(x_2) \cdot \dots \cdot P(x_7)}{P(x_1) + P(x_2) + \dots + P(x_7)}$$

Recordar que x es el vector que contiene como elementos el valor de cada una de las modalidades de la variable discriminante.

Fig. 20 Sentencias de programación cálculo probabilidad Bayesiana.

```

Wolfram Mathematica 6.0 [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb

0[x_, c_] := Sum[Coin[[c], i, x[[i]]]/b[[c]], {i, 1, 7}] (*dado c, 0 es la probabilidad condicional de que cliente x pertenezca a c.*)
Prob[c_, x_] := (0[x, c[[c]]) * b[[c]] / Sum(0[x, c[[i]]] * b[[i]], {i, 1, 7})) (*Probabilidad de que un individuo x pertenezca a un grupo c*)

```

Fuente: Elaboración propia.

3.3.5. Clasificación de Nuevos Clientes y Exportación de Resultados

Sólo resta ahora, programar las sentencias que permitan calcular a cada nuevo cliente su probabilidad de pagar, clasificarlos en pagadores y no pagadores por algún criterio, y por último, exportar los resultados en el sentido de extraer una base de datos con cada cliente, su probabilidad de pago y el grupo al cual pertenece.

Recordar que el objetivo es determinar cuáles de los nuevos clientes a facturar por primera vez su servicio, no pagaran su factura antes del vencimiento, por tanto, a cada nuevo cliente, el modelo calculará la probabilidad de pagar, y clasificará a todos los que tengan como probabilidad mayor a 0,4, respaldo por el lado de la empresa es reducir costos y centrar las acciones en los más “malos pagadores”) como pagador, caso contrario como no pagador, entonces para ello, se creará una primera variable, de nombre ?, esta variable almacenará para cada nuevo cliente el valor de su RUT, esto dado que será una información a exportar junto con su probabilidad y su clasificación en pagador o no pagador. Adicionalmente, se realizará un procedimiento el cual será nombrado como ?, que tendrá como objetivo realizar cuatro acciones, primeramente calcular la probabilidad de pagar de cada nuevo cliente, segundo comparar si está probabilidad es mayor que 0,4, si es así, lo clasificará como “Paga”, caso contrario como “No Paga”, y por último, exportará cada resultado a través de una matriz denominada ?, donde su primer elemento será el nombre de los datos a exportar, es decir, Rut, probabilidad, clasificación.

Una vez programado el procedimiento, se crea la variable ?, la cual será encargada de ejecutar el procedimiento teniendo como variable de entrada a ?, es decir, la matriz de nuevos clientes codificados y almacenará los resultados.

Para realizar la exportación de la clasificación, en *Wolfram Mathematica*, se utiliza la función:

Export["ruta", resultado, "Table", "TableForm"]

En la figura 22 se indican cada una de las sentencias programadas en el modelo para realizar la sección descrita.

Fig. 22 Sentencias para clasificación de nuevos clientes y exportación de resultados.

```

Wolfram Mathematica 7.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_CLASIFICACION.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_CLASIFICACION.nb
Wolfram Mathematica Player

Clasificando Nuevos Clientes

H = Z[[All, 7]];

Paga[Y_] := Module[{m, r, P},
  m = Length[Y];
  P = {"RUT", "PROB", "PAGA"};
  For[j = 1, j ≤ m, j++,
    r = Prob[1, Y[[j]]] // N;
    If[r > 0.4, AppendTo[P, {IntegerPart[H[[j]]], r, "Si"}], AppendTo[P, {IntegerPart[H[[j]]], r, "No"}]];
  P
];

PP = Paga[Z2];
PP[[11]]
  
```

Fuente: Elaboración propia.

Recordar que `H` contiene para cada nuevo cliente, su RUT, probabilidad de pagar y su respectiva clasificación en pagador o no pagador, ahora, para estimar que porcentaje de los nuevos clientes almacenados en `Z` son pagadores y no pagadores, se hace necesario primeramente programar una función capaz de contar los casos, para ello, `ContarPagadores` será la variable que contendrá cuantos clientes se clasificaron como pagadores, y `ContarNoPagadores` la variable que almacenará la cantidad de clientes clasificados como no pagadores. Para que tanto `ContarPagadores` y `ContarNoPagadores` obtengan sus respectivos valores, se les asigna la función predefinida de *Wolfram Mathematica* `Table` y `Count` y

?????? ???? ? ???? ???? respectivamente, es decir, función que cuenta los casos en la matriz clasificados como pagadores y no pagadores respectivamente. Con lo anterior, se posee la cantidad de clientes de cada grupo, al dividir el valor almacenado tanto en ? como en ? por la cantidad total de nuevos clientes, se está en condiciones entonces de determinar la razón de presencia pagadores y no pagadores del total de nuevos clientes.

Fig. 23 Sentencias de cálculo número nuevos clientes clasificados en pagadores y no pagadores.

```

Wolfram Mathematica 6.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_VF.nb

■ Calculo Pagadores y No Pagadores

G1 = Cases[PP, {___, "Si"}];
Length[G1] / Length[Z2] // N

0.173885

G2 = Cases[PP, {___, "No"}];
Length[G2] / Length[Z2] // N

0.826115

```

Fuente: Elaboración propia.

3.4. Testing

Se realiza el testing del modelo con los 2.000 clientes, a los cuales se les calcula su probabilidad de pagar, solicitando al modelo agrupar a todos aquellos con probabilidad de pago mayor a 0,4 en pagadores y a los que no en no pagadores, como se indicaba en el capítulo 3.3.5, la elección de este criterio se sustenta por un lado, desde la perspectiva empresarial, por la necesidad de reducir costos involucrados en la cobranza y centrar de esta forma los recursos en los más malos pagadores. Por otro lado, desde la perspectiva teórica la elección de este criterio de corte se sustenta por el

uso de una probabilidad a priori en base a un valor conocido previamente o por la hipótesis planteada.

Al realizar un cruce entre la clasificación de los clientes que conforman la base de testing y su real comportamiento se comprueba que el 58,23% de los clientes clasificados como no pagadores efectivamente tuvieron ese comportamiento al momento de pagar su primera boleta del servicio contratado con la empresa.

Fig. 24 Testing.

```
Wolfram Mathematica 7.0 - [Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_TESTING.nb *]
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
Modelo Clasificación Bayesiana_Morosidad Presunta Vencimiento_TESTING.nb *

Clasificando Nuevos Clientes TESTING

In[29]:= H = Z[[All, 7]];

In[30]:= Paga[Y_] := Module[{m, r, P},
  m = Length[Y];
  P = {"RUT", "PROB", "PAGA"};
  For[j = 1, j <= m, j++,
    r = Prob[1, Y[[j]]] // N;
    If[r > 0.4, AppendTo[P, {IntegerPart[H[[j]]], r, "SI"}], AppendTo[P, {IntegerPart[H[[j]]], r, "No"}]];
  P
];

In[31]:= PP = Paga[Z2];
PP[[11]]

Out[32]:= {153479895, 0.568019, SI}

■ Export Excel

In[33]:= Export["Pronostico TESTING2.xls", PP]

Out[33]:= Pronostico TESTING2.xls
```

Fuente: Elaboración propia.

4. RESULTADOS

De los 9.328 nuevos clientes, 3.293 fueron clasificados por el modelo como no pagadores de su primera factura al vencimiento, lo que representa un 35,3 % del universo, por otro lado, 6.035 de los clientes, el modelo los clasificó como pagadores, equivalente a un 64,7%.

Al revisar su estado de pago al día siguiente del vencimiento, de los 9.328 clientes, a 8.309 se les fue posible determinar el estado de su saldo, la diferencia se explica por clientes que dieron de baja el servicio durante el periodo de análisis de éste trabajo, y por tanto, se toma la decisión de excluirlos reduciéndose el universo de análisis a 8.309 nuevos clientes. De éstos, el 71,4% se encuentra aún sin pagar su factura al vencimiento, representando a 5.931 clientes.

Se posee hasta el momento, por un lado la clasificación de los clientes por el modelo en pagadores y no pagadores, y por otro, el estado de estos al día siguiente del vencimiento, por tanto, ahora corresponde realizar un cruce por rut para ver cuántos de los clientes que el modelo clasificó como morosos al vencimiento realmente quedaron en esa situación un día después del vencimiento de su factura.

El modelo clasificó como no pagador a 3.293 clientes, de los cuales 2.910 estaban activos y por tanto se les podría determinar su estado de morosidad al vencimiento, de estos, 2.179 estaban en mora al vencimiento, en conclusión, el 75% de los clientes clasificados por el modelo como no pagadores efectivamente quedaron en esa condición al día siguiente de haber vencido su primera boleta del servicio. Al exponer el resultado en la empresa, se ha estimado que es un modelo altamente valorable y por tanto será aplicado para los próximos vencimientos, es más, se desea replicar este modelo para el

servicio de televisión pagada satelital, cuya tecnología se conoce como DTH por sus siglas en inglés *Direct To Home* (*directa al hogar*).

Dado los favorables resultados del modelo, la gerencia corporativa de la empresa, ha solicitado a la jefatura de cobranzas, que redefina el modelo actual de “cobranza blanda”, en el sentido de redistribuir los esfuerzos y costos involucrados. Dado lo anterior, se ha determinado que el proceso de cobranza para los nuevos clientes siga la siguiente lógica operacional:

Paso 1, nuevos clientes a facturar: Todos los clientes instalados entre el día 16 del mes anterior y el día 15 del mes en curso, corresponderán a los clientes que por primera vez deberán pagar su factura del servicio contratado, siendo el date line de pago el día 4 del mes siguiente.

Paso 2, clasificar nuevos clientes: Entre el día 16 y 24 del mes en curso, se deberá preparar la base de clasificación, en el sentido que a cada cliente deberá confeccionarse su vector característico para acto seguido realizar su clasificación vía modelo.

Paso 3, monto a facturar nuevos clientes: El día 25 de cada mes, día posterior a la emisión de la factura de los clientes, se deberá determinar el monto a facturar a cada uno de los clientes que el modelo ha clasificado como presunto moroso al vencimiento.

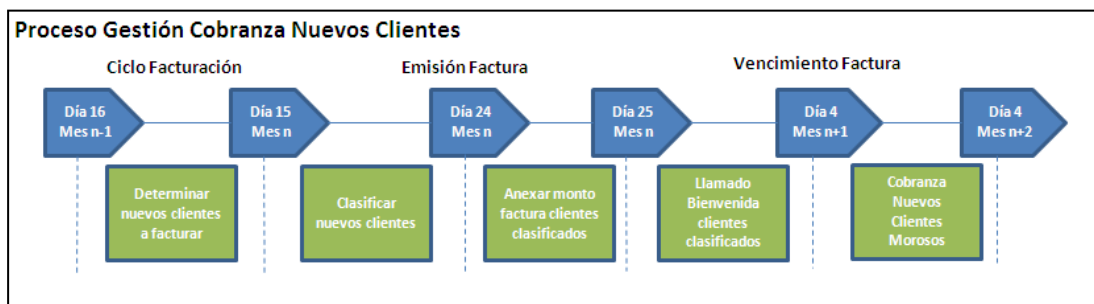
Paso 4, bienvenida de nuevos clientes: Entre el día 26 y el 4 del mes siguiente, se realizará una nueva acción, denominada como bienvenida, dónde vía callcenter se llamará a cada uno de los clientes clasificados como presuntos morosos, en este llamado se reforzara el servicio contratado, las características y valores de éste, como se compone su primera factura, las formas y lugares de pago del servicio, como también, se indagará si el servicio desde el punto de vista técnico está funcionando sin problemas, en caso de presentarse uno y al

considerarse como una variable de morosidad, estos casos serán transferidos en línea a la plataforma de retención de la empresa, dónde se buscará de manera inmediata dar solución al problema.

Paso 5, cobranza nuevos clientes: Todos aquellos nuevos clientes que no realizaron su pago al vencimiento, entrarán en un proceso de corte del servicio, el cual se realiza el día 7 de cada mes y entrarán a gestiones de cobranza, con ello se busca incentivar el pago total de lo adeudado, existiendo como alternativa la repactación de la deuda y el pago de una avance.

La figura 25 muestra el nuevo flujo de gestión de cobranza de la empresa para nuevos clientes anteriormente descrito.

Fig. 25 Nuevo flujo de gestión de cobranza.



Fuente: Empresa

5. CONCLUSIONES

Como se comentaba en la sección 4, del universo de nuevos clientes a clasificar por el modelo, 2.910 clientes éste los clasificó como no pagadores de su servicio al vencimiento de su factura, al revisar los saldos de estos clientes al día siguiente de su vencimiento, 2.179 de éstos efectivamente no habían realizado el pago aún, es decir, el 75% de los clientes clasificados como morosos efectivamente terminaron en esa condición. A raíz de los resultados obtenidos, me ha permitido validar la hipótesis de que los modelos de clasificación bayesianos aplicados sobre la morosidad presunta resultan útiles y efectivos para este tipo de aplicaciones, ahora bien, es necesario aclarar que al momento de realizar este proyecto existían algunas circunstancias inherentes propias al hecho de estar operando un negocio joven por parte de la empresa y que se intuía que podrían estar incidiendo en el no pago de un cliente de su factura, me refiero específicamente a la estabilidad de la red, sobre todo cuando un nodo ha sido liberado recientemente en un comuna, en el cual la señal recibida puede presentar algunas intermitencias, como también, el modulo de emisión de facturación del sistema que en el algunos casos ha emitido cobros inconsistentes, dado lo anterior, y considerando el equipo de trabajo desplegado por la empresa para disminuir estas incidencias en el menor plazo, se recomienda en el futuro realizar una nueva base de aprendizaje y realizar un nueva corrida de aprendizaje del modelo, calibración, y una nueva clasificación de clientes y validar nuevamente su resultado. Otras de las recomendaciones que pueden contribuir al perfeccionamiento del modelo pasan por el poder incorporar nuevas variables de ingreso asociada al cliente en el sistema al momento de contratar, por ejemplo, profesión u ocupación, edad, número de personas que aportan al ingreso familiar, cantidad de hijos, variables sociales que se caracterizan por estar asociadas a la responsabilidad de pago de un servicio independiente de cuál sea éste. Adicionalmente, se recomienda aplicar este modelo en los próximos ciclos de facturación, por ejemplo sobre clientes que enfrentan su segunda boleta, tercera, etc. Es más, se puede en una

primera instancia determinar los pagadores y no pagadores, y sobre este segundo universo clasificarlos según la acción de respuesta que estos tienen a distintos medios de cobranza, por ejemplo, llamado vía ejecutivo de call center, sms, carta o cobranza judicial. En fin, es posible dimensionar el abanico de posibilidades que estos modelos de clasificación pueden aportar en una empresa.

6. BIBLIOGRAFÍA

- [1] Albert, M; Aha D (1991). "Analyses of instances-based learning algorithms", Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), AAAI Press.
- [2] Abad, M. (2001). "Aplicación del principio inductivo de MEVR en la construcción de clasificadores", tesis doctoral. Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, España.
- [3] A. H. Guerra, Aprendizaje automático. Árboles de decisión, Veracruz: Universidad veracruzana. Pub., 2004.
- [4] Buntine,W.L. (1994) Operations for Learning with Graphical models. Journal of Artificial Intelligence Research, 2, 159-225.
- [5] Carroll, J.D.: Chang, J.J. (1970) "Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition", Psychometrika 35: 283-319.
- [6] Castillo, E; Álvarez E. (1997). "Sistemas Expertos. Aprendizaje e incertidumbre", Paraninfo, Madrid.
- [7] Castillo, W. (1991) "Descripción de algunos métodos de clasificación automática y aplicación a un problema de producción distribuida por cantón", Ciencias Matemáticas 2(1): 67-78.
- [8] Castillo, W.; González, J. (1994) "Análisis de tablas múltiples de datos", Revista de Matemática: Teoría y Aplicaciones 1(1): 47-55.

- [9] Chan, H. y Darwiche, A. (2004). Sensitivity Analysis in Bayesian Networks: From Single to Multiple Parameters. En Proceeding of the 20th Conference on Uncertainty in Artificial Intelligence, Arlington, Virginia, USA, pp. 67- 75. AUAI Press.
- [10] Chateld, C.; Collins, A.J. (1980) Introduction to Multivariate Analysis. Chapman & Hall, Londres.
- [11] C.J. Alonso, Técnicas de aprendizaje para la identificación de fallos en sistemas dinámicos, Valladolid: Grupo de sistemas inteligentes. Pub,. 2004.
- [12] Cooper, G. (1990). Computational complexity of probabilistic inference using Bayesian belief networks (Research note). Artificial Intelligence, 42, 393-405.
- [13] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. First International Workshop on Multiple Classifier Systems. I. J. K. a. F. Roli, New York: Springer Verlag. 1857: 1-15.
- [14] Ding, Q. and W. Perrizo (2002). Decision Tree Classification of Spatial Data Streams Using Peano Count Trees. Proc. of the ACM 124 Symposium on Applied Computing, Madrid, Spain.
- [15] González, J.; Rodríguez, O. (1995) "Algoritmo e implementación del método Stasis", en: J. Trejos (Ed.) IX Simposio de Métodos Matemáticos Aplicados a las Ciencias, U.C.R. - I.T.C.R., Turrialba.
- [16] Heckerman, D. (1995). A Tutorial on Learning With Bayesian Networks. Technical Report, Msr TR-95-06, Microsoft Research, Redmond, WA. Véase <ftp://ftp.research.microsoft.com>

- [17] Hernández, L.; Abad, M. (2001). "Clasificadores bayesianos robustos". Actas de la Novena Conferencia de la Asociación Española para la Inteligencia Artificial, Gijón, España, 2001.
- [18] Huete, J. F. (1995). "Aprendizaje de redes de creencia mediante la detección de independencias: Modelos no probabilísticos". Tesis Doctoral, Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada, España.
- [19] Kohavi, R. (1996). Scaling up the accuracy of naive-Bayes classifiers: a decision tree hybrid. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD96), Portland, OR., AAAI Press.
- [20] Lebart, L.; Morineau, A.; Fénelon, J.-P. (1985) Tratamiento Estadístico de Datos. Métodos y Programas. Marcombo, Barcelona.
- [21] Leigh, W; Purvis, R; Ragusa, J. (2002). "Forecasting the nyse composite index with technical analysis, pattern recognition, neural network, and genetic algorithm: a case study in romantic decision support", Decision Support Systems, 32, 361- 377.
- [22] L. M. Bergasa, Clasificadores estadísticos aplicados al reconocimiento de imágenes, Alcalá: Universidad de Alcalá. Pub., 2.004.
- [23] O. D. Castrillón, W.A. Sarache et all, "Algoritmo de Clasificación Bayesiano", 6ta Conferencia Iberoamericana en sistemas, Cibernética e Informática, Vol. 1, 2007, pp. 137 – 140.
- [24] Q. Cazorla. Miguel, Un enfoque bayesiano para la extracción de características y agrupamiento en visión artificial, Alicante: Universidad de alicante. Pub., 2000.

- [25] R. O. Duda, Pattern Classification, New York: Wiley, Interscience. Pub.,2002.
- [26] Shachter, R. y Kenley, C. (1989). Gaussian influence diagrams. Management Science, 35, 527-550.
- [27] Shachter, R. (1988). Probabilistic inference and influence diagrams. Operations Research, 36, 589-604.
- [28] Sahami, M. (1996). "Learning limited dependence Bayesian classifiers", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining", 335, 338, Menlo Park, CA.
- [29] Singh, M. and G. M. Provan (1995). Efficient Learning of Selective Bayesian Network Classifier. International Conference on Machine Learning. Philadelphia, PA., Computer and Information Science Department, University of Pennsylvania.
- [30] Yule, G.U. (1907). On the Theory of Correlation for Any Number of Variables, treated by a New System of Notation. Proc. Royal Soc. London, Serie A, 79, 182-193.

7. ANEXO I: CALCULO DE PROBABILIDADES

7.1. Acontecimientos y sus operaciones lógicas

7.1.1. Acontecimientos

Una prueba, o experiencia aleatoria, es una operación con resultado imprevisible, pero perteneciente a una lista bien definida Ω de resultados posibles; a Ω se le llama conjunto fundamental, espacio de los observables asociado a la experiencia.

Ejemplo, si la experiencia consiste en lanzar dos dados, su conjunto fundamental Ω son los 36 resultados siguientes, $\Omega = \{(1,1); (1,2); (1,3); (1,4); (1,5); (1,6); (2,1); (2,2); (2,3); (2,4); (2,5); (2,6); (3,1); (3,2); (3,3); (3,4); (3,5); (3,6); (4,1); (4,2); (4,3); (4,4); (4,5); (4,6); (5,1); (5,2); (5,3); (5,4); (5,5); (5,6); (6,1); (6,2); (6,3); (6,4); (6,5); (6,6)\}$, donde $??$, $?$ representa el número obtenido en el primer dado, $?$ el número obtenido en el segundo dado. A los elementos del conjunto fundamental se les denomina acontecimientos elementales de la experiencia. Al final de una prueba se puede preguntar por un acontecimiento específico, por ejemplo, la suma de los dados es 5. Un acontecimiento puede ser representado por un subconjunto de Ω , en el ejemplo:

$\Omega = \{(1,1); (1,2); (1,3); (1,4); (1,5); (1,6); (2,1); (2,2); (2,3); (2,4); (2,5); (2,6); (3,1); (3,2); (3,3); (3,4); (3,5); (3,6); (4,1); (4,2); (4,3); (4,4); (4,5); (4,6); (5,1); (5,2); (5,3); (5,4); (5,5); (5,6); (6,1); (6,2); (6,3); (6,4); (6,5); (6,6)\}$

7.1.2 Operaciones lógicas de acontecimiento

Sobre los acontecimientos, se definen operaciones lógicas tales como:

- **Igualdad de dos acontecimientos** $A = B$: Si siempre se realizan o no de manera simultánea, se describen como $A = B$.
- **Contrario de un acontecimiento** \bar{A} : Es el acontecimiento que se realiza o ocurre cuando A no lo hace y viceversa. Por lo tanto $\bar{\bar{A}} = A$ es representado por el complemento de A en Ω ; se escribe como: $\bar{A} = \Omega - A$.
- **Conjunción de dos acontecimientos** $A \cap B$: Es el acontecimiento que se realiza solamente si A y B se realizan simultáneamente. $A \cap B = A \cdot B$.
- **Reunión dos acontecimientos** $A \cup B$: Es el acontecimiento que se realiza siempre que A o los dos a la vez se hayan realizado. $A \cup B = A + B - A \cap B$.
- **Implicación:** Se dice que A implica B , si cada vez que A se realiza también lo hace B ; por tanto es $A \subseteq B$ una parte de B .

Adicionalmente se definen dos acontecimientos particulares: el acontecimiento **seguro** que se realiza siempre como resultado de la experiencia y que será representado por el conjunto Ω completo; y el acontecimiento **imposible**, que no se realiza jamás y que se representa por el conjunto vacío.

Si la conjunción de dos acontecimientos es imposible, se dice que son **incompatibles** $A \cap B = \emptyset$.

7.2. Probabilidad en un conjunto finito

Definición de probabilidad: Sea Ω un conjunto finito, y \mathcal{G} el conjunto de acontecimientos. La aplicación P será una probabilidad si es una función real definida para todo acontecimiento, y tal que:

- (1) La probabilidad de cualquier acontecimiento es no negativa. Para todo $A \in \mathcal{G}$, se tiene $P(A) \geq 0$.
- (2) El acontecimiento seguro tiene probabilidad 1, $P(\Omega) = 1$.
- (3) Si A y B son acontecimientos incompatibles, es decir, $A \cap B = \emptyset$, la probabilidad de la reunión es la suma de los acontecimientos: $P(A \cup B) = P(A) + P(B)$.

Una vez que ω haya sido escogida, se dice que se ha probabilizado el espacio de los acontecimientos, y todo lo que hay que conocer de la experiencia está entonces sintetizado en el triplete (Ω, \mathcal{G}, P) llamado **espacio de probabilidad finito**.

Propiedad 1

Sea el espacio de probabilidad finito (Ω, \mathcal{G}, P) y sean A y B dos acontecimientos cualesquiera $A, B \in \mathcal{G}$.

- (4) Si $A \subset B$, entonces $P(A) \leq P(B)$.
- (5) Para todo $A, B \in \mathcal{G}$, $P(A \cap B) \leq P(A)$ y $P(A \cap B) \leq P(B)$.
- (6) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(7) Si $P(A) = 0$, entonces $P(A \cap B) = 0$ para cualquier evento B ;

(8) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Una función de probabilidad definida para el conjunto fundamental finito Ω es conocida si se conocen todos los valores que toma para todos los acontecimientos, es decir para todos los subconjuntos $A \subseteq \Omega$. Se designará por $\omega_1, \omega_2, \dots, \omega_n$ los n acontecimientos elementales que constituyen los elementos de Ω ; se verá que en la práctica es suficiente conocer los valores correspondientes a los acontecimientos elementales $P(\omega_1), P(\omega_2), \dots, P(\omega_n)$ para definir P .

Propiedad 2

Sea el espacio de probabilidad finito (Ω, \mathcal{F}, P) y sean $\omega_1, \omega_2, \dots, \omega_n$ los n acontecimientos elementales. Se considera la aplicación real f definida en Ω por:

(9) $f(\omega_i) = x_i$.

Entonces el conocimiento de f permite calcular la probabilidad de cualquier acontecimiento $A \subseteq \Omega$ por:

(10) $P(f^{-1}(A)) = \sum_{\omega_i \in f^{-1}(A)} P(\omega_i)$.

En efecto, todo acontecimiento A se escribe como reunión de acontecimientos elementales ω_i que lo realizan; siendo dos a dos incompatibles, se tiene:

$$P(A) = \frac{n(A)}{n(S)} \quad P(B) = \frac{n(B)}{n(S)}$$

7.3. Equiprobabilidad y enumeración

7.3.1. Probabilidad Uniforme

Sea el espacio de probabilidad finito Ω con n acontecimientos elementales. Se dirá que la probabilidad es uniforme si todos los acontecimientos elementales tienen la misma probabilidad, es decir, $\frac{1}{n}$. Se dice también que los acontecimientos elementales son equiprobables.

En este caso la probabilidad $P(A)$ de un acontecimiento A puede evaluarse determinando el número $n(A)$ de acontecimientos elementales que realizan A .

$$(11) \quad P(A) = \frac{n(A)}{n}$$

Habitualmente se dice que la probabilidad es el cociente entre los casos favorables y el número de casos posibles.

Ejemplo 1: Una bolsa de dulces contiene seis mentas, cuatro chicles y tres chocolates. Si una persona hace una elección al azar de uno de los dulces, encontrar la probabilidad de sacar a) una menta b) un chicle o un chocolate.

Solución: $\omega_1, \omega_2, \dots, \omega_{13}$ representan los acontecimientos de que la persona seleccione, respectivamente, una menta, un chicle o un chocolate. El número total de dulces es 13, los cuales tienen la misma probabilidad de ser seleccionados.

$$(a) \quad P(A) = \frac{6}{13}$$

$$(b) \frac{?}{??}$$

7.3.2. Enumeración

En todos los casos los problemas de cálculo de probabilidades se reducen a problemas de enumeración (contaje del número de acontecimientos elementales que realizan, que se realizan más fácilmente por la aplicación de una serie de reglas, dentro de las cuales se verán a continuación las más relevantes.

Regla de multiplicación para elecciones sucesivas

Los trabajos que deben realizarse en dos etapas, donde la primera posee n maneras de realizar, y la segunda posee m maneras, en total habrá $n \cdot m$ maneras de realizar el trabajo. Si hay k etapas, y cada una puede realizarse de n_i maneras, en total habrá $n_1 \cdot n_2 \cdot \dots \cdot n_k$ maneras de realizar el trabajo.

Variaciones y permutaciones

Se tienen n objetos distintos, por ejemplo los enteros del $1, 2, \dots, n$. Sea r una lista ordenada de r objetos escogidos entre los n objetos iniciales es llamada ordenación de r objetos tomados de n en n . Se llamará permutación al caso que todos los objetos son ordenados n ; en el caso contrario, es decir, $r < n$, se llama variación de r objetos tomados de n en n . La construcción de una ordenación es un trabajo de r etapas con n posibles elecciones en la primera etapa, $n-1$ en la segunda, etc. El número de variaciones de r objetos tomados de n en n es:

$$V(n, r) = \frac{n!}{(n-r)!}$$

Combinaciones

Si el orden de extracción no se tiene en cuenta, el resultado de elegir r objetos entre n se le llama combinación de r objetos tomados de n en n . Se llama $\binom{n}{r}$ al número de combinaciones posibles.

$$(12) \quad \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Coeficientes multinomiales

Se consideran n objetos no todos distintos, de los cuales r_1 son rojos idénticos, r_2 son blancos idénticos; ...; r_k son negros idénticos, donde $r_1 + r_2 + \dots + r_k = n$. Si los n objetos fuesen todos distintos, habría $n!$ permutaciones; mientras que ahora hay $\frac{n!}{r_1! r_2! \dots r_k!}$ que son indiscernibles, $\frac{n!}{r_1! r_2! \dots r_k!}$ que también lo son, etc. Estas permutaciones ahora suman en total $\frac{n!}{r_1! r_2! \dots r_k!}$ entre las $n!$ totales.

El número de permutaciones distintas serán:

$$(13) \quad \frac{n!}{r_1! r_2! \dots r_k!}$$

A este número se le llama coeficiente multinomial.

7.4. Probabilidad condicional e independencia

Se sabe que un acontecimiento A se ha realizado, y se quiere tener en cuenta esta información para evaluar la probabilidad de realización de otro acontecimiento B . Después de n experiencias, se han observado r_1 realizaciones de A , y r_2 realizaciones simultáneas de A y B . La

frecuencias relativas de estos acontecimientos son $\frac{n_{11}}{n_{1.}}$, $\frac{n_{12}}{n_{1.}}$ y $\frac{n_{21}}{n_{2.}}$, $\frac{n_{22}}{n_{2.}}$ respectivamente. La frecuencia relativa de A_1 entre las veces que se ha realizado es:

$$\frac{n_{11}}{n_{1.}}$$

Lo cual sugiere la definición siguiente:

Definición de probabilidad condicional

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad finito; sean A_1 y A_2 dos acontecimientos con $P(A_1) > 0$. El cociente

$$(14) \quad \frac{P(A_1 \cap A_2)}{P(A_1)}$$

Se llama probabilidad condicional de A_2 sabiendo que A_1 se ha realizado.

El punto importante es que, para un A_1 fijo, la aplicación notada P_{A_1} que toma el valor $P_{A_1}(A_2)$ para todo acontecimiento A_2 de \mathcal{F} es una probabilidad, es decir, cumple con los tres axiomas:

- (a) La probabilidad de cualquier acontecimiento es no negativa. Para todo $A_2 \in \mathcal{F}$, se tiene $P_{A_1}(A_2) \geq 0$.
- (b) El acontecimiento seguro tiene probabilidad 1, $P_{A_1}(\Omega) = 1$.
- (c) Si A_2 y A_3 son acontecimientos incompatibles, es decir, $A_2 \cap A_3 = \emptyset$, la probabilidad de la reunión es la suma de los acontecimientos: $P_{A_1}(A_2 \cup A_3) = P_{A_1}(A_2) + P_{A_1}(A_3)$.

Propiedad 3

Sea Ω un espacio de probabilidad finito, y sea un acontecimiento tal que $A \subseteq \Omega$; entonces la aplicación P_A es tal que $(\Omega, \mathcal{G}, P_A)$ es un espacio de probabilidad.

Para todo $A \subseteq \Omega$, se tiene $P_A(A) = 1$; por otra parte, $P_A(A^c) = \frac{P(A^c)}{P(A)}$

Si $A \cap B \neq \emptyset$ se tiene $P_A(B|A) = \frac{P(A \cap B)}{P(A)}$, luego:

$$P(A \cap B) = P(A) P_A(B|A) = P(A) \frac{P(A \cap B)}{P(A)}$$

Si $A \cap B \neq \emptyset$ y $P(A) > 0$ es posible escribir las relaciones entre probabilidad y probabilidad condicional en una forma llamada ley de multiplicación:

$$(15) \quad P(A \cap B) = P(A) P(B|A)$$

Eliminando $P(A)$, se obtiene la relación llamada fórmula de Bayes:

$$(16) \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Ejemplo 2: La probabilidad de que un vuelo programado normalmente salga a tiempo es $P(A) = 0,82$; la probabilidad de que llegue a tiempo es $P(B) = 0,75$; y la probabilidad de que salga y llegue a tiempo es $P(A \cap B) = 0,65$. Encontrar la probabilidad de que un avión a) llegue a tiempo dado que salió a tiempo y b) salió a tiempo, dado que llegó a tiempo.

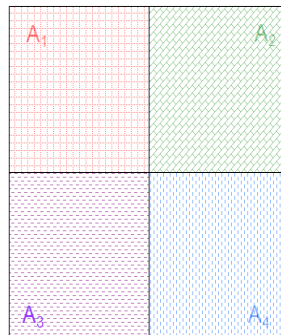
Solución

a) La probabilidad de que un avión llegue a tiempo, dado que salió a tiempo es

b) La probabilidad de que un avión saliera a tiempo, dado que llegó a tiempo es

A toda familia de acontecimientos que constituya una partición de Ω (son dos a dos disjuntos, y su reunión es Ω) se le llama una familia completa de acontecimientos.

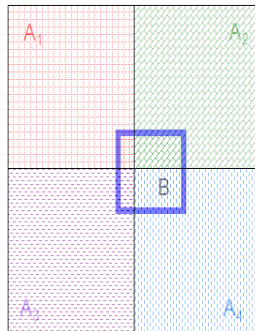
Fig. 26 Partición de Ω



Fuente: Elaboración propia

Todo acontecimiento puede escribirse como la reunión de A_i siendo los elementos de la reunión disjuntos;

Fig. 27 Descomposición de acontecimiento



Fuente: Elaboración propia.

Por consiguiente

Si $P(A_i|B) > 0$ para todo i , la ley de multiplicación conduce a la fórmula llamada ley de probabilidades totales:

(17)

Ejemplo 3: Cierta artículo es manufacturado por tres fábricas, F_1, F_2 y F_3 . Se sabe que la primera fábrica produce el doble que la segunda, y que la segunda fábrica produce igual que la tercera. Se sabe además que el 2% de los artículos producidos por las dos primeras fábricas son defectuosos, mientras que el 4% de los artículos producidos por la tercera fábrica son defectuosos.

Si se colocan todos los artículos en una fila y se escoge uno al azar ¿cuál es la probabilidad de que este artículo sea defectuoso?.

Solución: Sea S el espacio muestral completo y se consideran los siguientes eventos:

$$P(A_i) = \sum_{j=1}^n P(A_i | B_j) P(B_j)$$

Notar que $P(A_i) = \sum_{j=1}^n P(A_i | B_j) P(B_j)$, entonces:

$$P(A_i | B_j) = \frac{P(A_i \cap B_j)}{P(B_j)}$$

$$P(A_i \cap B_j) = \frac{P(A_i | B_j) P(B_j)}{P(A_i)}$$

Continuando, al aplicar la fórmula de Bayes al elemento B_j de la familia completa:

$$P(B_j | A_i) = \frac{P(A_i | B_j) P(B_j)}{P(A_i)}$$

Aplicando a B_j la ley de probabilidades totales (16) se obtiene el siguiente teorema:

7.5. Teorema De Bayes

Sea B_1, B_2, \dots, B_n una familia completa de acontecimientos tal que $B_i \cap B_j = \emptyset$ y sea A un acontecimiento tal que $A \cap B_j \neq \emptyset$. Entonces se verifica para todo j :

$$(17) \quad P(B_j | A) = \frac{P(A | B_j) P(B_j)}{\sum_{i=1}^n P(A | B_i) P(B_i)}$$

Interpretación: Imaginar que todos los acontecimientos B_j que particionan Ω puedan tener un efecto sobre el acontecimiento A , de manera que todo B_j pueda ser considerado como una posible causa del acontecimiento A . Entonces $P(B_j | A)$ se interpreta como la probabilidad de que B_j sea la causa de A ; el

teorema dice cómo puede ser calculada a posteriori la probabilidad de esta causa:

- a) Cuando se conocen las probabilidades a priori $P(A_i)$ de las causas.
- b) Como también las probabilidades condicionales $P(B|A_i)$ de realización del acontecimiento sabiendo que la causa A_i se ha realizado.

Puede ser que la realización de un acontecimiento no tenga ninguna influencia en la realización de . Esta posibilidad, muy importante en el razonamiento probabilístico y en sus consecuencias estadísticas, es formalizado de la siguiente forma.

Definición de acontecimientos independientes

Se dice que dos acontecimientos A y B son independientes si

$$(19) P(A \cap B) = P(A) \cdot P(B)$$

Esta definición se justifica por el hecho de que si A y B son independientes, se tiene $P(A|B) = P(A)$ que:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

De otra manera, el conocimiento de la realización de B no aporta ninguna información concerniente a la probabilidad de realización de A . Se debe tener en cuenta que si A y B son incompatibles, no significa que sean independientes, ya que entonces $P(A \cap B) = 0$ lo cual implica $P(A \cap B) \neq P(A) \cdot P(B)$, mientras que $P(A \cap B) = P(A) \cdot P(B)$ será en general no nulo. Igualmente si A implica B , $P(A \cap B) = P(A)$ luego $P(A \cap B) \neq P(A) \cdot P(B)$ $P(A \cap B) = P(A) \cdot P(B)$. Es fácil probar que si A y B son independientes, también lo son A y \bar{B} , \bar{A} y B así como \bar{A} y \bar{B} .

Ejemplo 4: Se sabe que el 30% de las lavadoras de cierta compañía requieren servicio mientras está vigente la garantía, en tanto que solo el 10% de sus secadoras necesitan este servicio. Si alguien compra una lavadora y una secadora de esta compañía, ¿cuál es la probabilidad de que ambas máquinas requieran servicio de garantía?

Solución: Sea A el acontecimiento en el cual la lavadora necesite servicio mientras está vigente la garantía y sea B el acontecimiento definido de manera análoga para la secadora. Entonces $P(A) = 0.30$ y $P(B) = 0.10$. Suponiendo que las dos máquinas funcionan de modo independiente, la probabilidad deseada es:

$$P(A \cap B) = P(A) \cdot P(B) = 0.30 \cdot 0.10 = 0.03$$

7.6. Ejemplo de Clasificación Bayesiana

Descrita la base probabilística para entender la clasificación bayesiana, en esta sección se buscará, a través de un ejemplo práctico, dar a conocer como se aplica la clasificación Bayesiana con el objeto de entender la lógica de programación del modelo.

Supongamos que se tiene el género y el puntaje de test psicológico para 15 trabajadores, además, se tiene una columna con la clasificación para cada individuo en:

```

?  ?? ??•??•??•?? ??????G
?  ???•? ??•??•??•?? ??????G
? ?  ???  ?????? ?????? ??????G

```

La variable género tiene dos modalidades ? ? ??? ????? ? ? ? ??????????, mientras que la variable test tiene 6 modalidades:

```

1 ?????????? ? ?? ????,
2 ???•???? ? ??? ????,
3 ?????????? ? ??? ????,
4 ?????????? ? ??? ????,
5 ?????????? ? ??? ????,
6 ?????????? ?????? ? ????.

```

La tabla siguiente, describe el vector característico de los 15 individuos a los cuales se les conoce ya su rendimiento, por tanto forman parte de lo que se conoce como la matriz de aprendizaje del modelo.

Nombre	Género	Puntaje	Rendimiento
Carolina	F	1	B
Eduardo	M	5	A
Andrea	F	4	M
Ximena	F	4	M
Consuelo	F	2	B
Matías	M	4	M
Lorena	F	1	B
Marcelo	M	2	B
Juan	M	6	A
Jorge	M	6	A
Mónica	F	3	M
Fernando	M	5	M
Paola	F	4	M
Paula	F	3	M
Verónica	F	3	M

Tabla1: base histórica ejemplo clasificación bayesiana.

Supongamos que se tiene una nueva fila de la base de datos (x, y) y se desea saber a partir de los datos históricos y usando clasificación bayesiana si Manuel corresponde a un individuo de bajo rendimiento, mediano rendimiento ó alto rendimiento laboral, es decir, si tiene mayor probabilidad de ser de bajo, mediano ó alto rendimiento.

Solución: Lo que se debe calcular es:

$P(B|1) = \frac{P(1|B) \cdot P(B)}{P(1|M) \cdot P(M) + P(1|B) \cdot P(B)}$

Se tiene de la tabla que:

$$P(1|M) = \frac{1}{17}, P(1|B) = \frac{1}{17}, P(M) = \frac{10}{17}, P(B) = \frac{7}{17}$$

Ahora bien, dado que el acontecimiento $A \cap B$ corresponde a dos acontecimientos independientes, ser Masculino y tener un puntaje clasificado como 5, aplicando independencia condicional se tiene que:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2}$$

$$P(A \cap B) = P(A|B) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$P(A \cap B) = P(B|A) \cdot P(A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Teniendo las probabilidades se está entonces en condiciones de calcular:

$P(A|B)$ y $P(B|A)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

$P(A|A)$ y $P(B|B)$

$$P(A|A) = \frac{P(A \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

$$P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

¿Por qué?

$$? \frac{? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ?}{? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ?$$

$$? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \frac{? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ?}{? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ? \text{ ? ? ? ? ? } ?}$$

Por lo tanto Manuel tiene mayor probabilidad de tener un alto rendimiento laboral.