



UTN.BA

FACULTAD
REGIONAL
BUENOS AIRES

TRABAJO FINAL INTEGRADOR

ESPECIALIZACIÓN EN INGENIERIA EN
SISTEMAS DE INFORMACIÓN

Título:

**Análisis comparativo de métodos en la predicción
de churn, en el ámbito de las telecomunicaciones**

Autor: Ing. Diego Hernán Feresin

Tutor: Dr. Hernán Merlino

Buenos Aires – Diciembre/2022

Resumen

En las organizaciones prestadoras de servicio los ingresos están dado por la cantidad de clientes que consumen sus servicios, la subsistencia consiste en el mantenimiento de dicha base de clientes, pero en mercados competitivos y saturados como el caso del mercado de las telecomunicaciones la adquisición de clientes sólo se logra por la captación de la competencia. De la misma forma la pérdida de clientes afecta a sus ingresos. En este caso poder prever la pérdida de estos clientes permite tomar acciones de retención y reducir la pérdida de ingresos. A partir del análisis de la información que se puede recoger de los distintos sistemas que poseen las compañías recuperando patrones permite realizar predicción del comportamiento de los clientes. Existen distintos estudios que utilizan una variedad de técnicas de predicción y diferentes técnicas de evaluación sin haber una que sea común o sobresalga sobre todos. Las fuentes de datos y los contextos tampoco son similares, por lo que para un contexto específico lo necesario será encontrar la mejor combinación que consiga los mejores resultados.

Abstract

In service-providing organizations, income is given by the number of clients that consume their services, subsistence consists of maintaining the client base, but in competitive and saturated markets, such as the telecommunications market, the acquisition of clients it is only achieved by attracting from the competition. In the same way, the loss of customers affects income. In this case, being able to predict the loss of these clients allows taking retention actions and reduce the income loss. From the analysis of the information that can be recovered from the different information systems the companies have, it is possible recognize patterns that allow predicting customer behavior. There are different studies that use a variety of prediction techniques and different evaluation techniques without one that is common or stands out above all. Data sources and contexts are not similar either, so for a specific context it is necessary to find the best combination that achieves the best results.

Tabla de contenido

Capítulo 1	1
1.1 Introducción	1
1.2 Objetivo	2
1.3 Alcance	2
1.4 Fundamentos del trabajo	2
1.5 Estructura del trabajo	4
Capítulo 2	4
2.1 Estado de la cuestión	4
2.2 Desarrollo del mapeo sistemático de la literatura	5
2.2.1 Planificación	5
2.2.2 Realización de la revisión sistemática	5
2.2.3 Generación de resultados.....	6
2.2.4 Reportes del resultado del SMS.....	6
Capítulo 3	24
3.1 Análisis y puntos relevantes del estado del arte	24
Capítulo 4	27
4.1 Conclusiones	28
4.1.1 Resumen de los resultados del trabajo.....	28
4.1.2 Futuras líneas de investigación.....	28
4.2 Pruebas de concepto	29
Referencias	32
Apéndice A: Listado de estudios primarios utilizados en el SMS	35
Anexos	39
Glosario de métodos y menciones	39
Atributos identificados	43
Datos Gráfico 16. Métodos de predicción mencionados.....	55
Gráfico 17. Métodos utilizados en los artículos.....	56

Índice de tablas

Tabla 1. Enlace de inicio de las búsquedas	5
Tabla 2. Repositorios de ubicación de artículos	6
Tabla 3. Repositorios artículos.....	7
Tabla 4. Publicaciones por años.....	8
Tabla 5 Cantidad de artículos por publicaciones	9
Tabla 6. Publicaciones con más artículos de las referencias	12
Tabla 7. Participación de autores.....	14
Tabla 8. Referencias de publicaciones	15
Tabla 9. Clasificación según Umayaparvathi.....	16
Tabla 10. Clasificación según Khan	16
Tabla 11. Clasificación según Mahajan y Ismail	16
Tabla 12. Categorías según Ismail	16
Tabla 13. Etiqueta 1 sobre atributos	17
Tabla 14. Etiqueta 2 sobre atributos	17
Tabla 15. Atributos clasificación 1.....	17
Tabla 16. Atributos clasificación 2.....	18

Índice de gráficos

Gráfico 1. Penetración telefonía móvil.....	3
Gráfico 2 Portabilidad telefonía móvil.....	3
Gráfico 3. Repositorios artículos	7
Gráfico 4. Distribución publicaciones.....	8
Gráfico 5. Artículos por palabras de búsqueda	8
Gráfico 6. Lugar de publicación.....	10
Gráfico 7 Lugar de publicaciones de referencias.....	11
Gráfico 8. Publicaciones y referencias	12
Gráfico 9. Autores.....	12
Gráfico 10. Origen Investigadores	13
Gráfico 11. Atributos por rango de dataset.....	18
Gráfico 12. Cantidad de registros por dataset.....	19
Gráfico 13. Herramientas para métricas.....	19
Gráfico 14. Métodos ajuste desbalanceo	20
Gráfico 15. Distribución de mención de métodos por año.....	21
Gráfico 16. Métodos de predicción mencionados	22
Gráfico 17. Métodos utilizados en los artículos	23
Gráfico 18. Métodos de reducción de dimensiones de los dataset.....	23
Gráfico 19. Cantidad de SMS en los últimos años.....	27

Capítulo 1

En este capítulo se encuentra la introducción del trabajo, en la sección 2 los objetivos, en la 3 el alcance del trabajo, en la sección 4 fundamentos del trabajo y en la sección 5 la estructura del trabajo.

1.1 Introducción

En el ámbito de las empresas de servicio, los ingresos están dado por las ventas que realizan a sus clientes. Para servicios del tipo cautivos o monopólicos la base de clientes se mantiene estable, como el servicio de distribución eléctrica, distribución de red de agua potable, etc. Para el caso de servicios en competencia, no sólo las empresas tienen una parte del mercado, si no que para aumentar sus ingresos deben aumentar el ingreso por cliente (ARPU del inglés Average Revenue Per User – Promedio de Ingreso Por Usuario) o realizar esfuerzos para incrementar esta base de clientes.

De la misma forma que pueden tener mayor penetración en el mercado, los clientes pueden tomar la decisión de abandonar el servicio, este abandono del servicio se conoce también como CHURN. En (Jain et al., 2021) propone una clasificación, que corresponde principalmente al área de las telecomunicaciones, para entender las posibles razones de este abandono. Lo clasifican en **No Intencional** e **Intencional o Voluntario** y dentro del **Intencional/Voluntario**, subclasifica en **Incidental** y **Deliberado**.

El caso del *No Intencional* es originado por la compañía, donde cortan el servicio como por ejemplo la falta de pago. En el caso del *Voluntario* la baja se da por el traslado del cliente a una zona geográfica sin cobertura de servicio o el cambio de trabajo donde la compañía tenga contrato con otro prestador. Y por último el *Deliberado* donde el usuario no está satisfecho con el servicio por diferentes razones, (Ullah et al., 2019) calidad del servicio, problemas en la facturación, tecnología entre otros, esto le permite al cliente probar otras compañías. Una de las razones que prevenía esta migración era la pérdida del número telefónico y el inconveniente de tener que distribuir el nuevo número a los contactos, pero a partir de la implementación de la portabilidad numérica (cambio de compañía telefónica y manteniendo el mismo número de abonado) esta barrera se redujo. De estas clasificaciones sobre el único tipo que pueden trabajar corresponde al *Deliberado*. Es por ello que han desarrollado estudios sobre la predicción del CHURN (Ullah et al., 2019)(Kostić et al., 2020)(Huang et al., 2015)(Azeem et al., 2017). Pero no es el único sector que trabaja sobre el abandono, también se da en el área de Finanzas, Juegos, Seguro, Servicios de noticias , Retail, etc. (Ahn et al., 2020). Especialmente en el ámbito de las telecomunicaciones las empresas de telefonía en conjunto con la predicción del churn, desarrollan programas de retención, esto previene la disminución de la cantidad de clientes, ya que compensar la

pérdida con la captura de nuevos clientes resulta más costoso (Amin et al., 2017) En Argentina , no sólo es un área de servicios muy competitivo sino que es un mercado saturado con una penetración de más del 100 % (Penetración nacional de la telefonía móvil (accesos por cada 100 habitantes) · Datos Abiertos ENACOM, s. f.). desde antes del año 2013, por lo que la implementación de esta predicción de abandono sumado a programas de retención resulta económicamente favorable.

Será necesario realizar una revisión de la situación en el dominio de las telecomunicaciones, determinar que métodos se utilizan para poder predecir el abandono. Cuáles son los datos que se toman como fuentes y determinando el conjunto de clientes con posibles intenciones de abandono, identificar sobre que característica sería necesario aplicar las acciones de retención

1.2 Objetivo

Objetivo principal del trabajo:

- Poder identificar los métodos utilizados para la predicción del churn en el área de las telecomunicaciones

Objetivos secundarios:

- Determinar los atributos utilizados para la predicción del churn
- Identificar las herramientas de evaluación de los métodos de predicción

1.3 Alcance

El alcance del presente trabajo estará enfocado en la determinación de las técnicas utilizadas para la predicción del CHURN en el área de las telecomunicaciones.

Para ello se realizará una revisión sistemática de la literatura para determinar el estado del arte en esta área. Se analizará los artículos para responder los objetivos planteados. Se presentará las conclusiones y se establecerán las próximas líneas de investigación.

1.4 Fundamentos del trabajo

Para todas las áreas de servicio los clientes son la fuente de los ingresos. En una etapa inicial de desarrollo de producto los objetivos están centrados en el crecimiento de esta base de clientes.

Con el transcurrir del tiempo y en base al éxito del servicio, la curva de crecimiento podrá ser mayor o menor, pero finalmente llegará a cubrir el mercado potencial donde ese crecimiento se hace nulo. Si bien el mercado tiene un volumen , la cantidad de ese mercado a la que puede acceder la organización dependerá de cuantos jugadores operen sobre ese conjunto de clientes.

En particular el mercado de la telefonía móvil en la Argentina se encuentra saturado desde hace varios años.

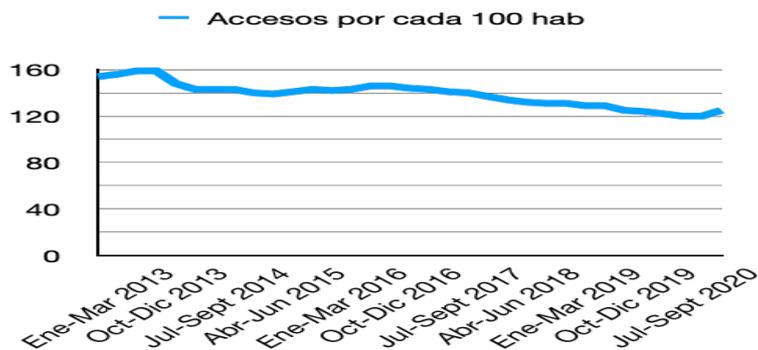


Gráfico 1. Penetración telefonía móvil¹

En estas condiciones los incrementos de los ingresos se dan por la prestación de más servicios, incremento de los valores de los servicios, crecimiento de clientes por la captación de la competencia.

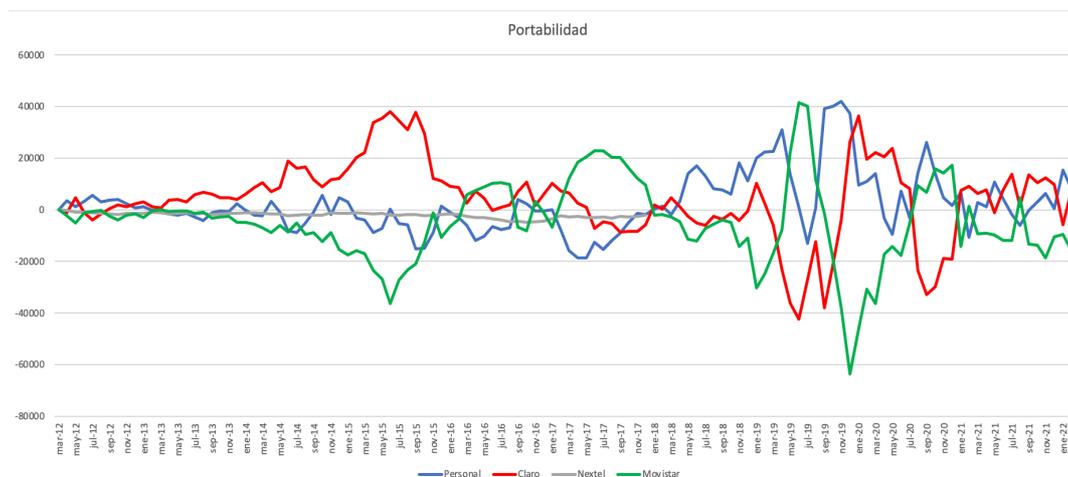


Gráfico 2 Portabilidad telefonía móvil²

Como puede verse en el Gráfico 2 se representa las migraciones de clientes entre las compañías, al ser un mercado saturado, las picos de uno de los prestadores se compensa con los valles de las otras. Estas variaciones se dan por campañas que se dedican a la captura de clientes. Por otro lado el abandono del cliente puede darse por razones de aumentos de costos, baja calidad de servicio, pérdida de funcionalidades (Amin et al.,

¹ Fuente <https://datosabiertos.enacom.gob.ar/dashboards/19998/telefonía-movil/>

² Fuente: (Portabilidad numérica móvil: portaciones netas mensuales por operador · Datos Abiertos ENACOM, s. f.)

2019), y de una forma pasiva la pérdida de una operadora se transforma en ganancia para otra, produciendo el efecto no deseado en la primera de reducción de ingresos. Para tratar de evitar este efecto no deseado, en las compañías existen equipos de retención que se encargan de tratar de convencer a los usuarios cuando se comunican para solicitar la baja del del servicio. Esta acción es reactiva y deben convencer al cliente para rever la decisión. La alternativa es la de poder trabajar en forma proactiva y adelantarse a la intención de los clientes del abandono también proponiendo incentivos pero en principio sobre una decisión no tomada. El éxito de estas acciones permitiría reducir la pérdida en los ingresos. En (Idris et al., 2019) indica que los costos de la captación de un cliente de la competencia es entre 5 y 6 veces mayor que lo que se dedica a los planes de retención. Por lo tanto, poder identificar los métodos utilizados para la detección anticipada de la posible decisión de abandono de cliente y aplicar los planes de retención es económicamente atractivo para las empresas de telecomunicaciones.

1.5 Estructura del trabajo

En el capítulo 1 se presenta la introducción del planteamiento del trabajo y la estructura.

En el capítulo 2 se desarrolla el estado de la cuestión en base un mapeo sistemático de la literatura (en Inglés Systematic Mapping Studies o SMS) tomando como guía el documento de Kitchenham. Se realizará la presentación de los datos de los artículos

En el capítulo 3 se realizará el análisis de los datos orientado a contestar las preguntas planteadas en el capítulo 2

Por último, en el capítulo 4 se presentarán las conclusiones del trabajo y futuras líneas de investigación.

Capítulo 2

En este capítulo se encontrará 1 el estado de la cuestión y en 2 el desarrollo del mapeo sistemático de la literatura

2.1 Estado de la cuestión

Para la realizar el estado de la cuestión se tomó como guía el documento de Kitchenham (Kitchenham & Charters, s. f.)

En primer lugar, se formularon las preguntas para determinar una guía para la investigación

P1: Cuales son las técnicas utilizadas para la detección de churn

P2: Cuales son los atributos utilizados para la detección de churn

A continuación, se definieron lo que serán los criterios de aceptación de los artículos que surjan

CA1: Publicaciones dentro de los últimos 8 años

CA2: El área de las publicaciones relacionadas con las telecomunicaciones

CA3: Revisiones que incluyan dentro de las áreas de investigación telecomunicaciones.

También se definieron los criterios de rechazo

CR1: Artículos no relacionados con el área de las telecomunicaciones

En base a las preguntas y a los criterios de aceptación se definieron las palabras a utilizar para las búsquedas

churn prediction

telecom "churn prediction"

"churn prediction"

2.2 Desarrollo del mapeo sistemático de la literatura

En esta sección se encontrará en la sección 1 la planificación del mapeo, en la 2 el proceso de revisión, en la sección 3 los resultados del proceso y en la sección 4 el reporte de los resultados.

2.2.1 Planificación

Una vez definido las preguntas de investigación, los criterios de aceptación y rechazo, las palabras claves, se procederá al descubrimiento de los artículos a través de los motores de búsqueda especializados en la actividad académica.

Se establece una plantilla básica para ir catalogando los artículos que respondan a los criterios de búsqueda.

A continuación, se analizarán los artículos tabulando la información disponible que sirva para responder las preguntas de investigación

Se conformarán los datos estadísticos a partir de la información tabulada.

Se procederá con un análisis de la información recolectada, finalizando con las conclusiones de trabajo y las futuras líneas de investigación

2.2.2 Realización de la revisión sistemática

Con los criterios definidos se procedió a realizar la búsqueda en internet, se utilizó como base de búsqueda de documentos.

Fuente	URL
google scholar:	https://scholar.google.com

Tabla 1. Enlace de inicio de las búsquedas

A partir de la aplicación de las palabras definidas para la búsqueda, se encontraron enlaces en otros repositorios y accesos directo que se detallan a continuación.

Fuente	URL
IEEE:	https://ieeexplore.ieee.org
ScienceDirect Elsevier:	https://www.sciencedirect.com/
SPRINGER:	https://link.springer.com/
ACM DL:	https://dl.acm.org/
Cornel University:	https://arxiv.org
HRCAK:	https://hrcak.srce.hr
International Journal of Scientific Research and Engineering Development	http://www.ijared.com/
Global Vision Press	https://gvpress.com/
International Research Journal of Engineering and Technology (IRJET)	https://www.irjet.net/
MDPI:	https://www.mdpi.com/
Researchgate:	https://www.researchgate.net

Tabla 2. Repositorios de ubicación de artículos

Con los artículos que cumplían con los criterios se analizaron y presentaron las características en Reportes del resultado del SMS.

2.2.3 Generación de resultados

Para las estadísticas de iniciales se procesó con sólo la plantilla y las funciones básicas de las hojas de cálculo. Para el caso de la información relacionada con los autores, las referencias y las publicaciones fue necesario incluir la misma en una base de datos debido a la cantidad y los cruces necesarios para poder extraer la información. Las dificultades correspondieron a la carga y la estandarización de los datos, por ejemplo los nombres de los autores figuraban en algunos casos como “inicial. apellido” y en otras “apellido, inicial”, esto impedía que la comparación directa produjera resultados . Fue necesario comparar, ordenar, ajustar los datos, simplificar los duplicados y reemplazar las referencias, los nombres originales están asociados a las publicaciones y las referencias, al eliminar un duplicado fue necesario reemplazar el valor del vínculo. Situaciones similares sucedieron con el nombre de las publicaciones de referencia y lugares de publicación.

2.2.4 Reportes del resultado del SMS

En base a los análisis de los artículos se pueden desarrollar las siguientes estadísticas.

Con las cadenas de búsquedas se encontraron 52 artículos luego de la aplicación de los criterios de rechazo se redujo a 47.

Los números obtenidos del resultado de las búsquedas fueron los siguientes.

Los artículos según las fuentes presentan la siguiente distribución

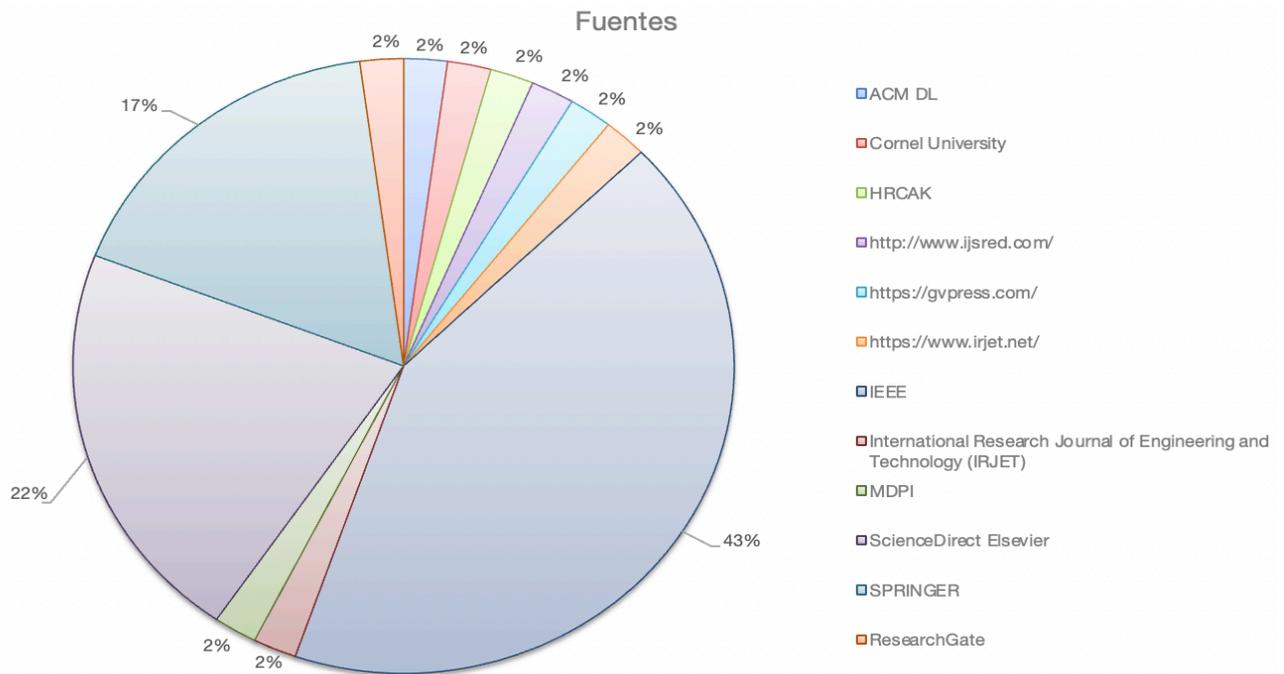


Gráfico 3. Repositorios artículos

Repositorio	cantidad
IEEE	20
ScienceDirect Elsevier	10
SPRINGER	8
ACM DL	1
Cornel University	1
HRC AK	1
http://www.ijsred.com/	1
https://gvpress.com/	1
https://www.irjet.net/	1
International Research Journal of Engineering and Technology (IRJET)	1
MDPI	1
ResearchGate	1

Tabla 3. Repositorios artículos

La distribución de los años de publicación se encuentra en el Gráfico 4



Gráfico 4. Distribución publicaciones

Año	Publicaciones
2015	6
2016	7
2017	16
2018	4
2019	7
2020	7

Tabla 4. Publicaciones por años

La cantidad de artículos por las palabras de búsqueda fueron los siguientes



Gráfico 5. Artículos por palabras de búsqueda

Los artículos fueron publicados en congresos, conferencias o journals cuya distribución se puede observar en el Gráfico 6.

Las publicaciones con más artículos del SMS son Cluster computing, IEE Access, International Research Journal of Engineering and Technology (IRJET), Telecommunication Systems y European Journal of Operational Research.

Para el caso de las referencias de los 1683 publicaciones se pudieron identificar 1334 (79%) , para los cuales del 50 % (lugares de publicación que tuvieron hasta 6 artículos) la distribución es la que se muestra en el Gráfico 7

El total de la distribución de la publicaciones de las referencias se repartió como se indica en la Tabla 5

Cantidad de artículos	Cantidad de publicaciones	Cantidad de artículos	Cantidad de publicaciones
368	1	1	13
75	2	1	14
24	3	1	16
4	4	3	17
6	5	2	18
6	6	1	19
8	7	1	21
4	8	1	22
3	9	1	23
2	10	1	28
2	11	1	30
2	12	1	208

Tabla 5 Cantidad de artículos por publicaciones

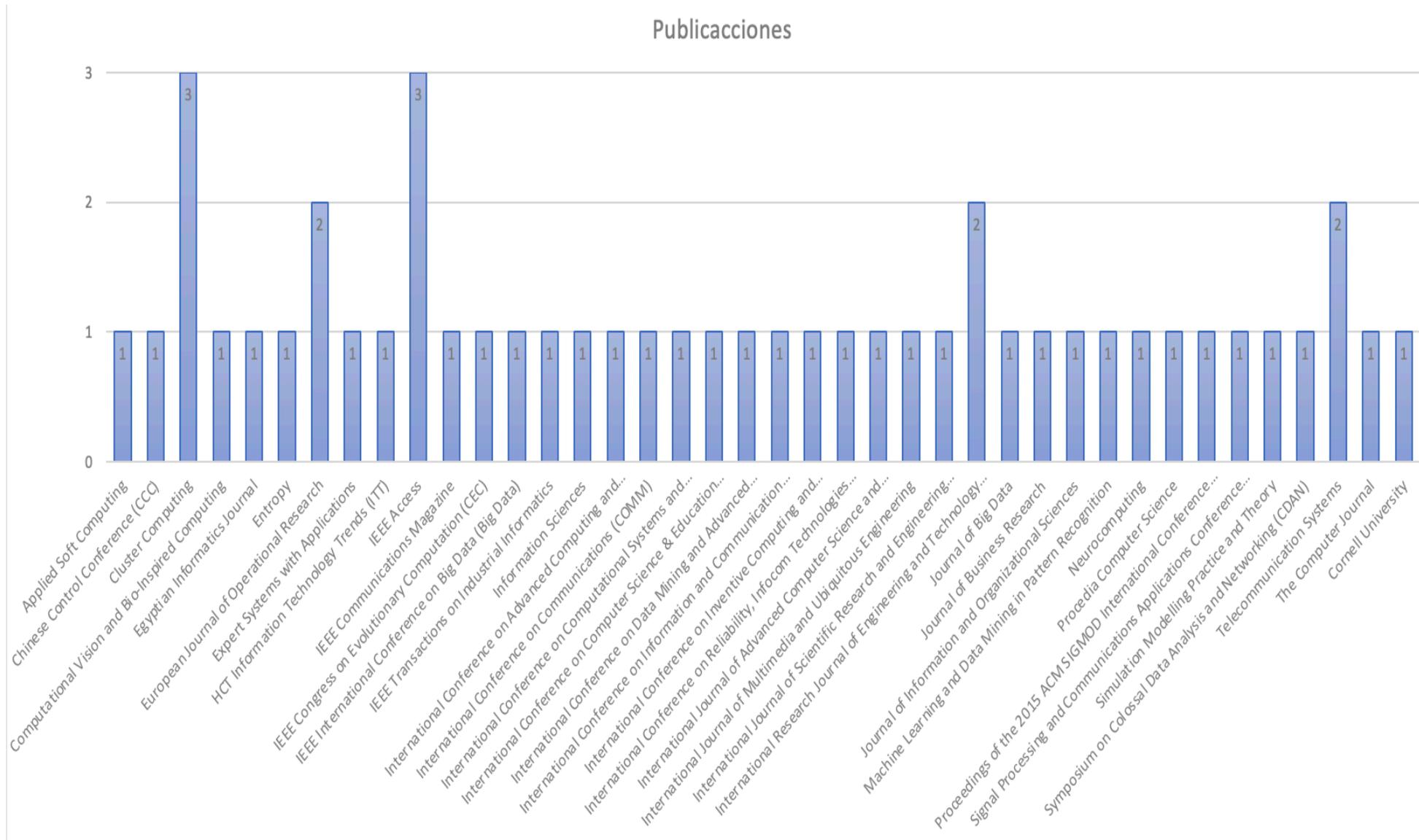


Gráfico 6. Lugar de publicación

Lugares de Publicaciones y referencias

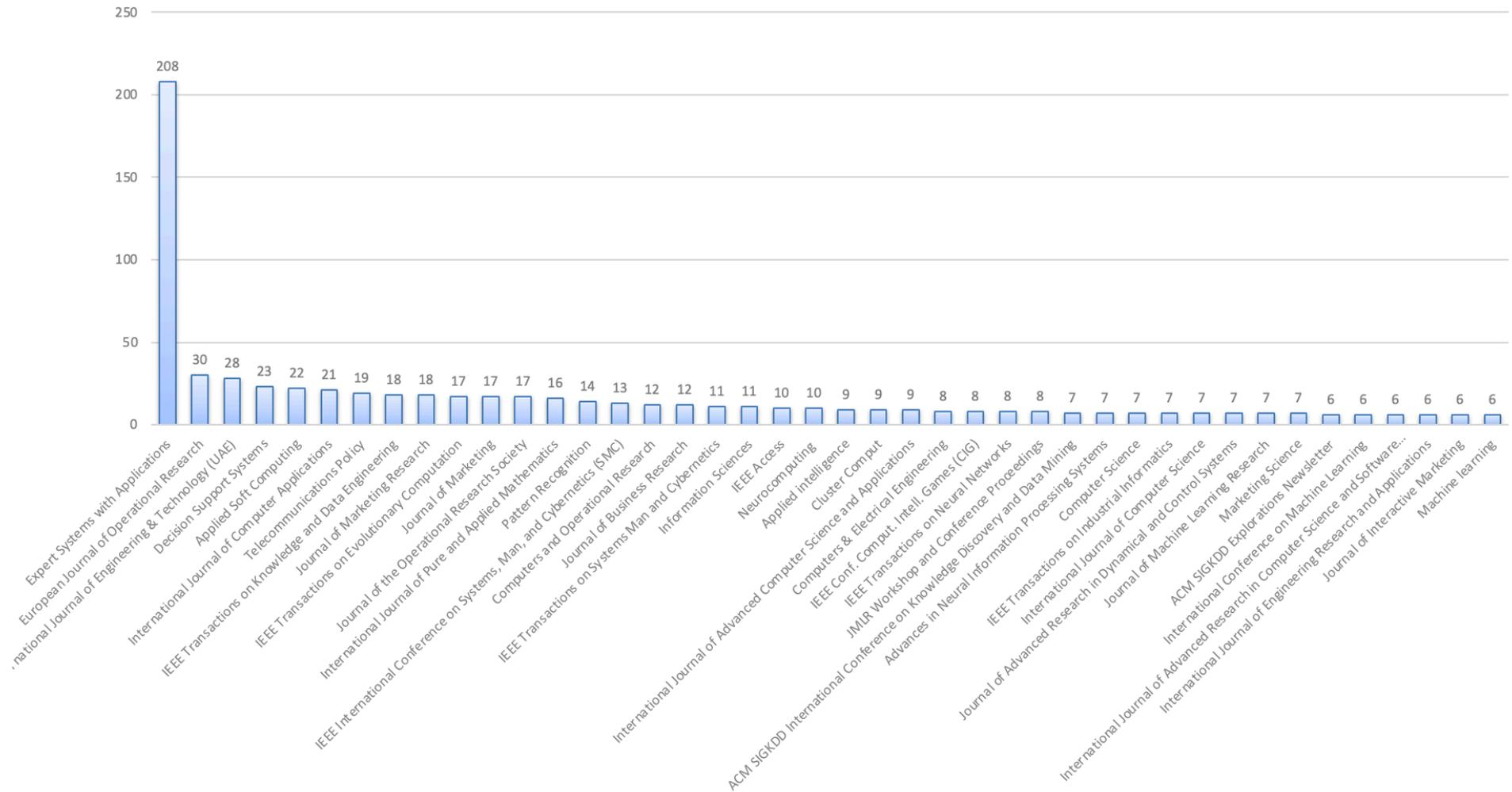


Gráfico 7 Lugar de publicaciones de referencias

Se ordenaron las publicaciones con más artículos de las referencias y el listado se encuentran en la Tabla 6

Publicación
Expert Systems with Applications
European Journal of Operational Research
International Journal of Engineering & Technology
Decision Support Systems
Applied Soft Computing
International Journal of Computer Applications
Telecommunications Policy
IEEE Transactions on Evolutionary Computation
IEEE Transactions on Knowledge and Data Engineering
Journal of Marketing Research

Tabla 6. Publicaciones con más artículos de las referencias

El conjunto de artículos hace referencia a 1683 publicaciones y dentro de esas referencias se encuentran 25 en las publicaciones utilizadas para el mapeo sistemático

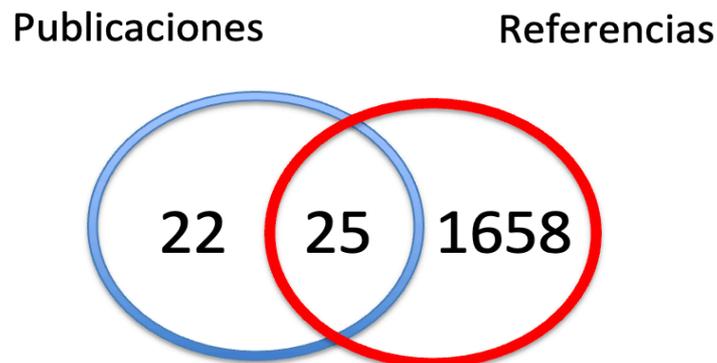


Gráfico 8. Publicaciones y referencias

Pasando al análisis de los autores, en las 47 publicaciones colaboran 122 autores, 3 autores en 3 publicaciones, 8 en 2 publicaciones y 111 autores sólo en 1 de 40 publicaciones en colaboración.

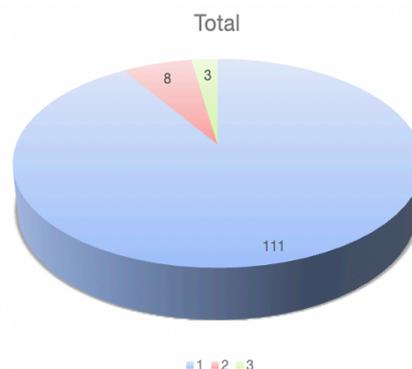


Gráfico 9. Autores

Los autores de los artículos pertenecen a distintos centros de investigación , en 11 artículos los autores pertenecen a centros de diferentes países y la distribución de países es la siguiente.

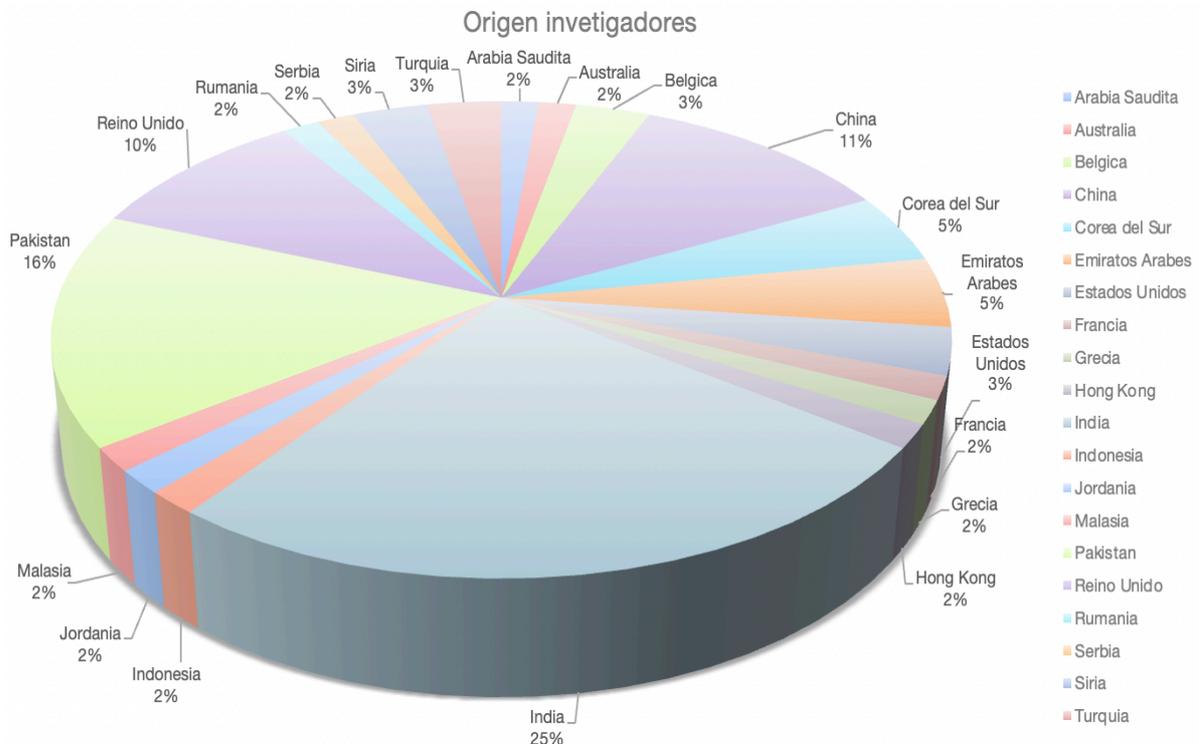


Gráfico 10. Origen Investigadores

2289 autores conforman el conjunto que colaboran en las publicaciones, siendo B. Baesens, D. Van den Poel, A. Maselena ,W. Verbeke, D. Martens los más prolíficos en cuanto a participación en publicaciones dentro de las referencias, siendo estos los primeros en la tabla de autores participantes.

La producción de los autores recogido en las referencias de los artículos las podemos describir de la siguiente manera.

Cantidad de autores	Cantidad artículos en que participan
1	82
1	71
1	62
1	48
1	39
1	32
1	30
1	29
5	28
2	25
1	24
2	22
1	21
1	20
1	19
2	18
1	17
1	16
3	15
8	14
4	13
13	12
5	11
10	10
12	9
18	8
34	7
43	6
55	5
97	4
158	3
347	2
1457	1

Tabla 7. Participación de autores

75 autores del conjunto de publicaciones de la revisión sistemática tienen publicaciones también dentro de las referencias.

La relación de las publicaciones y las referencias pueden representarse en la Tabla 8, de las 1683 referencias 871 son referenciadas por 1 publicación del SMS , 130 son referenciadas por 2 artículos , y desde el otro lado de la tabla, hay 16 artículos del SMS que hacen referencia a un única publicación, 15 artículos también tienen en común una publicación, etc.

Cantidad de publicaciones referenciadas	Cantidad de artículos de SMS que referencia a publicaciones
871	1
130	2
44	3
25	4
15	6
11	5
4	7
3	9
2	10
2	11
1	8
1	12
1	13
1	14
1	15
1	16

Tabla 8. Referencias de publicaciones

En lo referente a los objetivos secundarios, de los 47 artículos se pudieron obtener información de atributos utilizados para el análisis de CHURN en 26 de ellos. Los atributos de los cuales se tomó la información no se corresponden con la totalidad de los conjuntos de datos ya que en unos artículos se presentan los atributos más relevantes luego de procesos de priorización y como se puede ver en el Gráfico 11 la cantidad indicada en uno de los artículos es mayor que el total de los recolectados.

Se identificaron 528 atributos diferentes de los 785 recuperados de los artículos, siendo el identificador de CHURN, la permanencia, la edad y el género los atributos que más veces aparecen en los distintos conjuntos de datos.

Uno de los artículos (Ahmad et al., 2019) utiliza información del tipo de redes sociales, identificando amigos considerando la posible influencia que podría tener en la decisión de abandono del servicio en el resto de su entorno.

Los atributos en su mayoría se recuperaron de tablas dentro de los artículos, sin embargo cada uno de ellos tienen un origen, en algunos casos varios atributos provienen de la misma fuente, por lo que se los puede agrupar dentro de una categoría.

En sus artículos Umayaparvathi (Umayaparvathi & Iyakutti, s. f.-b) (Umayaparvathi & Iyakutti, 2016) (Umayaparvathi & Iyakutti, s. f.-a) utiliza la clasificación de los atributos en 6 categorías ,

Clasificación de atributos
Detalle de servicio de atención al cliente
Detalles personales y demográficos del cliente
Puntaje de crédito del cliente
Detalle de pagos y facturas
Patrones de uso del cliente
Servicios de valor añadido del cliente

Tabla 9. Clasificación según Umayaparvathi

Khan en cambio (Khan & Kozat, 2017) utiliza una clasificación referenciando al tipo de datos

Clasificación por tipo
Catagóricos
Numéricos estacionarios
Numéricos secuenciales
Tipo de salida

Tabla 10. Clasificación según Khan

En tanto V. Mahajan (Mahajan et al., 2015) como M. R. Ismail (Ismail et al., 2015) clasifican los atributos por el tipo de información similar a como lo hace Umayaparvathi

Mahajan clasifica en

Tipo de información
Información de facturación de utilización del cliente
Comportamientos de uso
Estado de churn de usuarios pasados y tipos de plan.

Tabla 11. Clasificación según Mahajan y Ismail

Ismail lo hace en las categorías de

Tipo de información
Demografía de cliente
Información de relacionamiento del cliente
Información de facturación
Información de uso

Tabla 12. Categorías según Ismail

Estas clasificaciones se tomaron como base para la agrupación de los atributos encontrados en conjunto de datos particulares. Para poder hacer una valoración de los atributos totales utilizados se incorporaron 2 clasificaciones para poder agrupar el tipo de atributo utilizado.

La primera clasificación se hizo en base a la descripción del atributo o lo que representaba el atributo medido como lo hizo Khan.

En este sentido a cada atributo se le asignó una de las siguientes etiquetas

Etiqueta 1
Datos
Días
Estático
Fecha
Llamadas
Mixto
MMS
SMS
Tiempo
Variable

Tabla 13. Etiqueta 1 sobre atributos

La segunda clasificación se hizo en base al tipo de valor que representaba el atributo.

Etiqueta 2
Cantidad
Costos
Duración
Valor Medio
Minutos
Valor

Tabla 14. Etiqueta 2 sobre atributos

Los atributos quedaron agrupados de la siguiente forma

Según descripción	Cantidad de atributos
Datos	23
días	18
estático	77
fecha	2
llamadas	242
mixto	50
MMS	10
SMS	50
tiempo	18
variable	295

Tabla 15. Atributos clasificación 1

Según el tipo de valores de los atributos

Tipo de valor	Cantidad de atributos
cantidad	207
costos	140
duración	63
medio	39
minutos	81
valor	255

Tabla 16. Atributos clasificación 2

Esto nos permitirá a posteriori evaluar el tipo de atributos sobre los cuales trabajar.

Este conjunto de atributos se extrajo de la información recopilada de los artículos, y como se mencionó anteriormente, no representa la totalidad de los datos que se toman como fuente para los procesos de predicción, mientras que en sólo 26 se expresaban los atributos de manera parcial en una tabla, en 32 artículos se identifican la cantidad de registros de 59 dataset, y en 25 artículos se identifican la cantidad de atributos de 52 dataset.

En Gráfico 11, se puede ver la agrupación de la cantidad de atributos por rango informados en los artículos

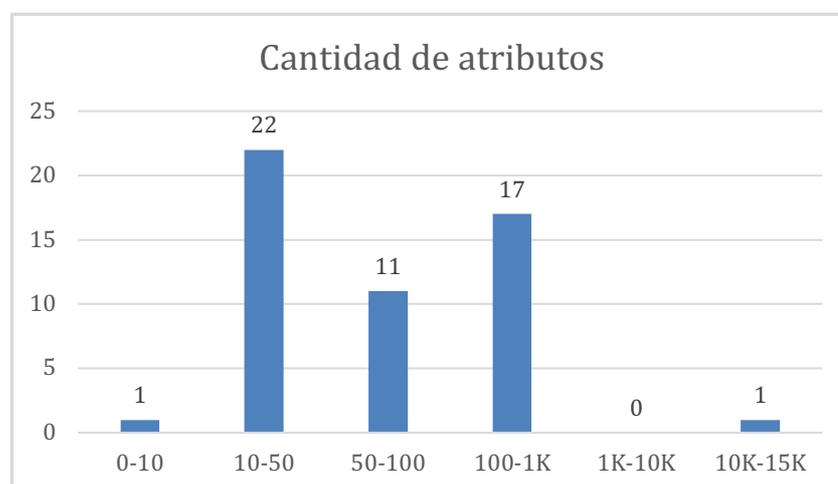


Gráfico 11. Atributos por rango de dataset

También para tener en cuenta es la cantidad de registros sobre los cuales se hacen las predicciones. Algunos artículos realizaban trabajos sobre dataset públicos alocados en universidades y otros en base a información de operadoras telefónicas, los casos de los 2 y 5 millones de registro los referencian a operadoras de China y Siria.

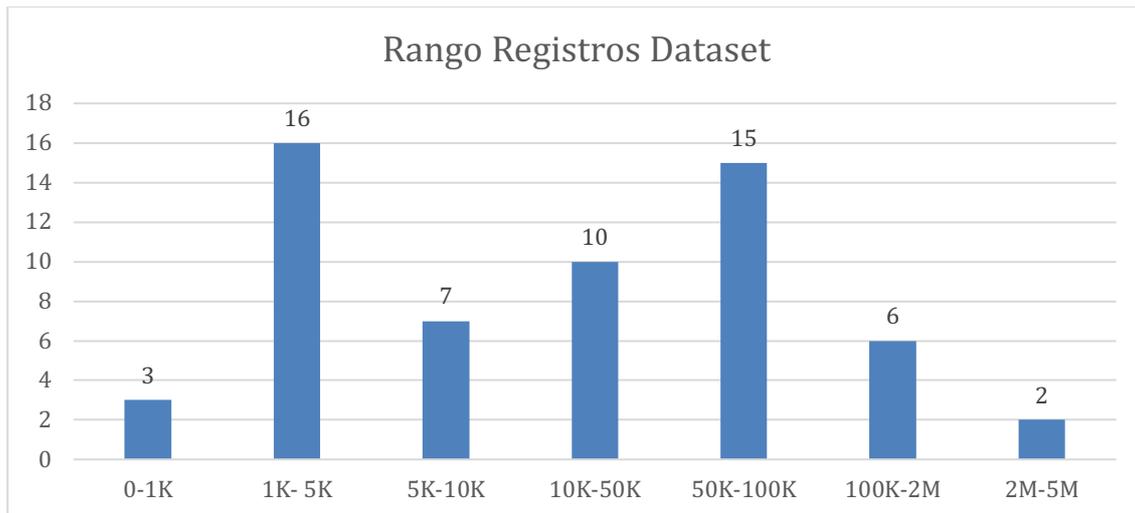


Gráfico 12. Cantidad de registros por dataset

Una vez analizado las características de los dataset, cantidad de atributos y cantidad de registros, otro de los objetivos secundarios a conseguir consiste en identificar cuáles son los métodos de evaluación de los modelos utilizados en la predicción.

De la revisión de las publicaciones surgieron los datos con los que se pueden armar el Gráfico 13 que representa el uso de las distintas herramientas para la medición y comparación de los de los métodos aplicados para la predicción.

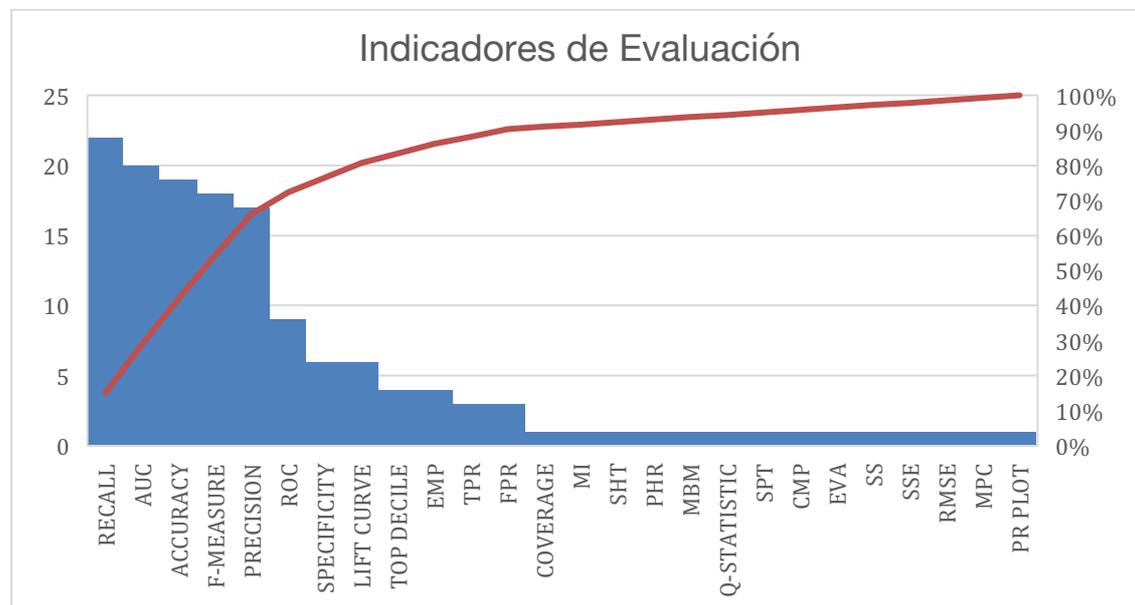


Gráfico 13. Herramientas para métricas

Dentro de las primeras herramientas , Recall, Accuracy, Precision, Specificity se calculan a partir del procesamiento de la matriz de confusión, F-Measure a partir de Precision y Recall (ambos funciones de la matriz de confusión), mientras que AUC (del inglés Area Under the Curve, area bajo la curva) , la curva es la generada por el valor de ROC.

Si bien están identificados de forma diferente Recall y Specificity es la misma métrica. Se mantienen en forma diferenciada para respetar la identificación en los artículos.

Con la fuente que son los datos y las herramientas para la medición y comparación, lo siguiente es continuar por los métodos. Sin embargo estos métodos pueden no ser efectivos por las proporciones que existen en los dataset entre los conjuntos mayoritarios y minoritarios. Algunos dataset que se encuentran en repositorios están preparados para realizar esas investigaciones donde la proporción entre los clientes clasificados como churn, y los no churn es del 50%. En el área de telecomunicaciones los clientes marcados como churn puede ser entre el 20 y 40% anual (Hassonah et al., 2019) esto representa entre 1,6% y 3,3 % mensual. En un caso de telefónica móvil en China (Huang et al., 2015) diferencia los datos de los prepagos con una media del 9,4% frente a los postpagos (con contratos) con una media del 5,2%. Estas diferencias tan importantes dificultan a las herramientas para poder detectar los clientes indicados como churn.

Para poder compensar este problema, se pre-procesan los dataset para modificar la proporción de las muestras. En un caso la estrategia es de sobre-muestreo, uno de los algoritmos selecciona elementos vecinos de las clases minoritaria cercana y genera nuevas muestras basada en la interpolación lineal (Synthetic Minority Oversampling Technique o SMOTE) aumentando la clase minoritaria. La segunda estrategia es eliminando en forma aleatoria muestras de las clases mayoritaria (Random Under-Sampling o RUS), este proceso tiene como desventaja la eliminación de muestras potencialmente útiles del dataset. (Zhu et al., 2017).

Dentro de los artículos se encontraron los siguientes métodos de modificación de las proporciones de muestreo Gráfico 14

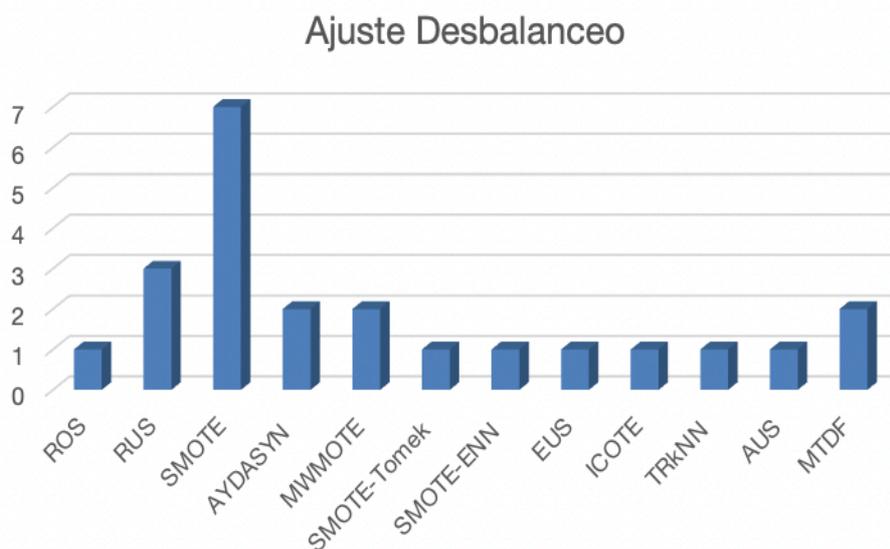


Gráfico 14. Métodos ajuste desbalanceo

El método SMOTE es el más referenciado y podríamos sumar las otras versiones SMOTE-Tomek y SMOTE-ENN incrementando la diferencia con las otras alternativas.

Para responder al último de los objetivos planteados del trabajo, en los artículos se han aplicado distintos procedimientos para la predicción, dentro de los artículos se han mencionado o utilizado métodos simples y combinados.

El método más mencionado corresponde a las Redes Neuronales. Le siguen los Árboles de Decisión, Regresión Lineal, SVM (Support Vector Machine), Random Forest, Naive Bayes, KNN (K-Nearest Neighbor), Genetic Algorithm, K-MC (K-Means Clustering). Estos 9 métodos representan el 50% de las menciones que se hacen respecto del uso para la predicción del churn. Y el otro 50% está representado por 129 métodos, con un total de 138 identificados.

La distribución de las menciones de los métodos por año se puede ver en el

Gráfico 15, para el año 2017 se encuentran la mayor cantidad de artículos de esta recopilación por lo que es lógica una mayor concentración en ese año.

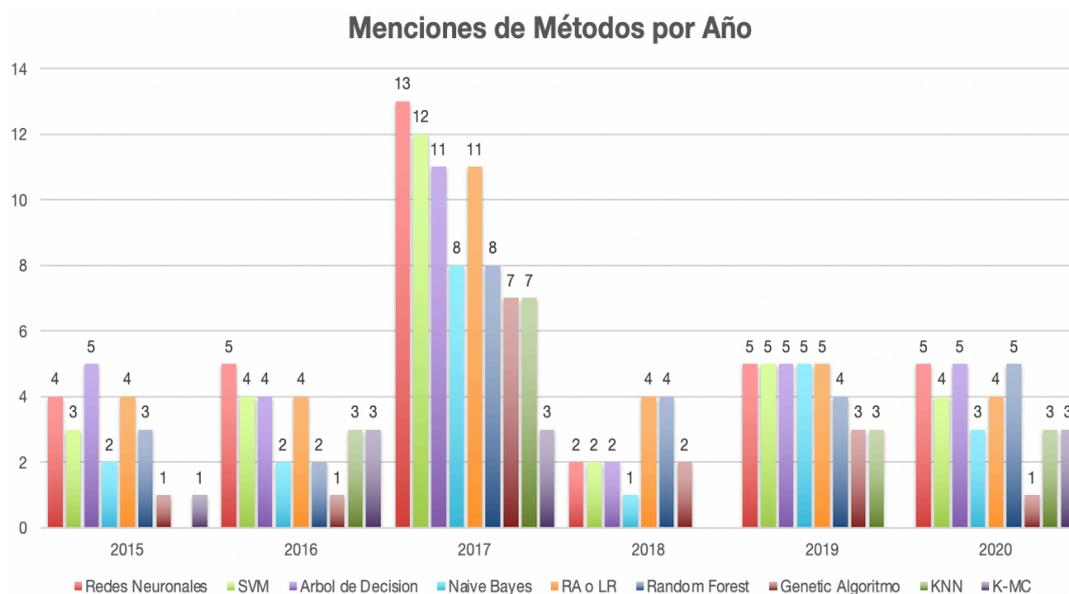


Gráfico 15. Distribución de mención de métodos por año

La distribución de los procedimientos identificados en los artículos se representa en el Gráfico 16

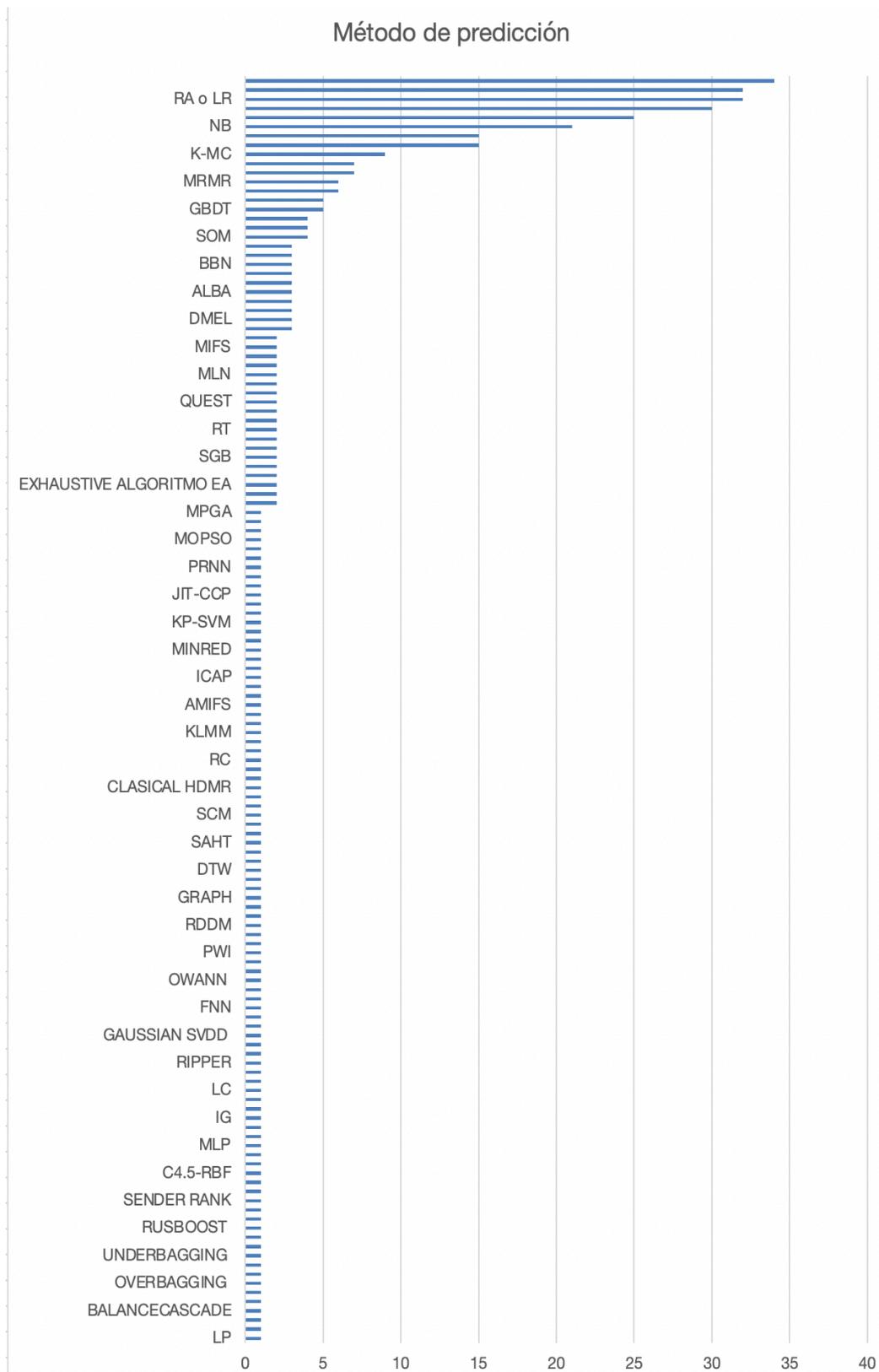


Gráfico 16. Métodos de predicción mencionados

Si bien los precedentes son métodos que se mencionan en los artículos, excluyendo los artículos que responden a revisiones, los algoritmos que se utilizaron o sirvieron de

comparación para mostrar las ventajas o mejoras respecto de las del tipo testigos se indica la siguiente distribución del

Gráfico 17

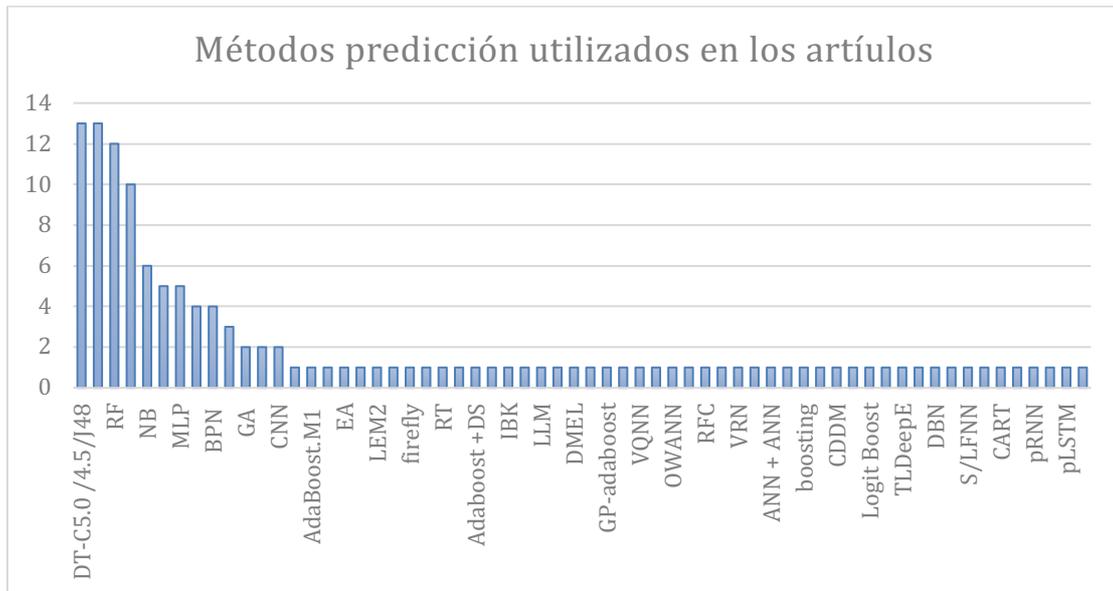


Gráfico 17. Métodos utilizados en los artículos

Otro de los procedimientos aplicados sobre los conjuntos de datos consiste en reducir las dimensiones, pues no todos los atributos tienen la misma influencia en el resultado de la predicción y la dimensión del conjunto de datos respecto a la cantidad de atributos puede influenciar en el tiempo de entrenamiento de los modelos, para esto se utilizan algoritmos para priorizar/evaluar el peso de cada uno de los atributos y de esta forma reducir el volumen de datos a computar.

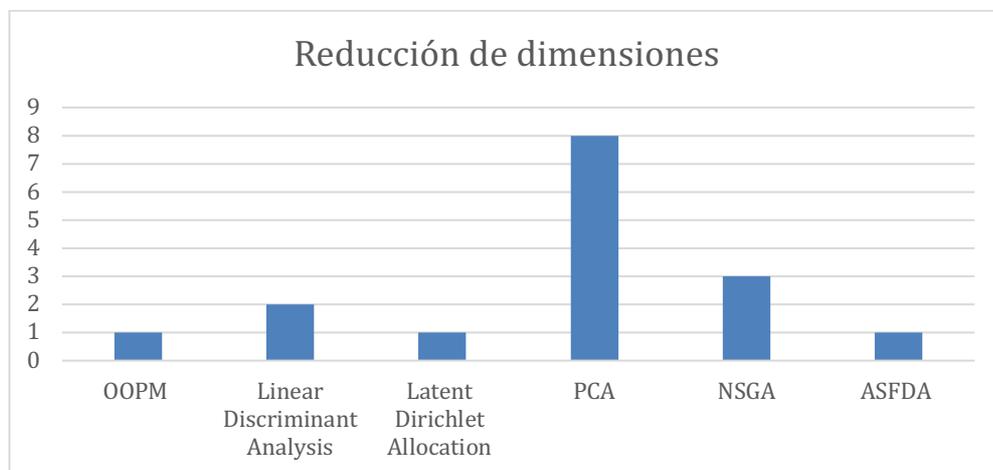


Gráfico 18. Métodos de reducción de dimensiones de los dataset

Si bien hay artículos que no identifican un procedimiento, se encontraron en forma explícita (Gráfico 18) los métodos en las proporciones indicadas

Capítulo 3

En este capítulo se desarrolla el análisis de los puntos relevantes del estado del arte

3.1 Análisis y puntos relevantes del estado del arte

En base a los criterios de aceptación y rechazo todos los artículos seleccionados tratan a la predicción del churn pero su contenido puede dividirse en distintos tópicos.

Artículos que son revisiones del tema donde cuentan con documentos relacionados con las telecomunicaciones además de incorporar otras áreas.

Un segundo grupo donde los artículos comparan procedimientos o presentan algún método utilizando otros conocidos para establecer comparaciones y las mejoras de predicción

Otro grupo presenta información sobre implementaciones en operadoras

Para los casos analizados se presenta una gran dispersión entre los conjuntos de datos que se utilizan y que han podido identificarse. Si bien son ejemplos de datos de empresas de telecomunicaciones, el 86% de los dataset contiene menos de 100.000 elementos (42% más de 10.000 y 44% menos de 10.000 muestras) por lo que los resultados podrían ser representativos en base a las proporciones analizadas pero podría ser un problema la aplicación práctica de la utilización de la información de las empresas, un caso que hace notar (Mitrović et al., 2018) en su artículo.

Los atributos identificados se relacionan principalmente con información de uso necesario para la facturación y características de los usuarios. Uno de los artículos que presenta un caso práctico , incorpora además información de la red utilizando software y hardware especializado para el manejo de información (Huang et al., 2015) , mientras que el volumen de la información de los CDR (detalle de llamadas del ingles Call Detail Record) es de 24Gb por día, la información que recolectan de la red es de 2,2 Tb. La incorporación de esta información incrementa los resultados, justificando la inversión por la cantidad de clientes que maneja la empresa (el estudio corresponde a uno de los casos de 2 millones de clientes de la muestra) y la posibilidad de retención de un gran número de ellos.

El uso de las redes neuronales es uno de los métodos más referenciado en los artículos, pero todas las reglas de decisión se encuentran dentro de la misma herramienta siendo una “caja negra” en la forma que predice el churn del cliente. Esto puede ser una desventaja en la aplicación práctica del uso de la detección del abandono, la utilización de otros métodos además genera información de las razones por las cuales determinados clientes tienen posibilidad de abandono, de esta forma se pueden aplicar políticas de retención más precisas para un grupo de clientes específico (Dalvi et al., 2016) (Ullah et al., 2019) (Amin et al., 2017)

Este concepto de segmentación lo han aplicado también a nivel de predicción. En primer lugar realizan una segmentación de los clientes en base a las características y luego a cada segmento aplican el método sobre un esquema de datos más acotado o con características más similares lo que produce un mejor resultado (De Caigny et al., 2018)

El otro caso de segmentación es aplicado al ARPU (ingreso promedio por usuario del inglés Average Revenue Per User) al conjunto de datos, ya que la principal razón es la de conservar los clientes por temas económicos, y desde este punto de vista, los que más ingresos proporcionan son más valiosos para las operadoras, el artículo pone de manifiesto que si bien la predicción es un proceso de clasificación, en la mayoría de los casos no hacen discriminación entre los costos del error de clasificar mal a un cliente (todos los clientes valen lo mismo) y que en la práctica no es el caso. (Wang et al., 2017) También dentro de la idea de una visión de negocio uno de los métodos de evaluación utilizado se centra en la ganancia que produce los clientes Expected Profit Measure (EMP) mencionados en (Zhu et al., 2017)(Alboukaey et al., 2020)(Amin et al., 2018)(Hu et al., 2018).

En los casos de utilización de métodos del tipo redes neuronales, la cantidad de atributos son una preocupación desde el punto de vista computacional, es por eso que aplican métodos para reducir la dimensión de los dataset sólo utilizar los atributos que tengan más influencia en la definición del abandono, técnicas como PCA. En contraposición a esto, en un artículo aplican métodos de Deep Learning, como la red neuronal convolucional que una de sus aplicaciones es la de reconocimiento de imágenes, en este caso la imagen estaría compuesta por la gran cantidad de los atributos de los dataset, convierten el vector de datos en una matriz asimilando a un formato de imagen y la clasificación la realizan a partir del análisis de las matrices. Los resultados no son mejores que otro artículo del mismo autor, pero es una estrategia novedosa para el manejo de volumen de atributos. (Umayaparvathi & Iyakutti, s. f.-c)(Ahmed et al., 2019).

También relacionado con las dimensiones, en 6 artículos utilizan información relacionada con datos de varios meses, en general corresponden a información de un periodo definido y los valores a los acumulados de ese período, en estos casos fragmentan los valores de los atributos. Mientras que en unos casos las muestras representan una “foto” de los patrones del cliente, al incorporar relaciones con otros “momentos” podría llegar a ser más representativo de un comportamiento.

Para el caso de los atributos que se han podido identificar, según la agrupación que se ha realizado en relación con los valores de estos, la máxima agrupación corresponde a atributos que tienen un **valor**, en este conjunto se agrupan atributos del tipo texto o que pueden mapearse en algún código como género, edad, plan internacional, código de área. También en este grupo se encuentra el indicador de churn. El segundo conjunto hace referencia a cantidades del tipo de valores numéricos en su mayoría del tipo entero, la palabra **cantidad** figura en su descripción (en la traducción de los nombres de los atributos) cantidad de llamadas, cantidad de llamadas entrantes, cantidad de llamadas de voz, cantidad de SMS y otros valores numéricos total de llamadas internacionales etc. En tercer lugar, atributos que representan **costos**, variables del tipo económicas, cargos totales, cargos del día, de la tarde, de noche, costos totales diarios etc. El siguiente grupo referencia atributos que tiene que ver con las llamadas expresado en **minutos**, entrantes, locales, internacionales, libres, total del día. El quinto conjunto que tiene que ver con la **duración** de las llamadas que no las identifican en minutos de la agrupación anterior, (22 de los 41 atributos de este conjunto tienen como descripción “duración de llamada”), también el tiempo de permanencia es un atributo que aparece en 13 ocasiones en los distintos dataset. Por último, los atributos que hacen referencia a promedios o valores **medios** como promedio de llamadas, promedio de llamadas internacionales, promedio de llamadas locales, etc.

La siguiente clasificación que se ha hecho ha sido en base a la descripción del mismo atributo, en primer lugar se han agrupado los que no entraba en el resto de las clasificaciones donde se le agregó el rótulo de **variable**, de alguna forma los valores de los atributos pueden modificarse a lo largo del tiempo, los atributos en esta categoría pueden ser número de mensajes, monto de pago promedio, facturación mensual, amigos fuera de la red, etc. . el siguiente en cantidad es el rotulo llamadas y aquí encontramos: cantidad de llamadas en el día, duración de la llamada, duración de las llamadas entrantes, duración de las llamadas locales, llamadas fuera de la red etc.. La siguiente agrupación es de los atributos estáticos como género, ciudad, educación, código de área, id cliente, etc. Si bien alguno de los atributos puede cambiar a lo largo del tiempo, son más estables en comparación de los variables. La siguiente clasificación mixto agrupa principalmente a los

cargos globales como cargos totales del día, cargos de noche, costo total diario, costo total por mes. La clasificación de mixto se interpretó por la agrupación de distintas características dentro de este atributo. Son pocos atributos en esta clasificación, pero se han repetido en varios conjuntos de datos identificados, 16 atributos y 50 instancias. Con la misma cantidad de instancias la clasificación que identifica los SMS. Cantidad de SMS, local, internacional, diario, plan de SMS, etc. **Datos** es el siguiente rótulo identificado, en este grupo encontramos, velocidad de internet, uso de datos, paquetes de bajadas, paquetes de datos comprados, etc.. Dos grupos también con la misma cantidad de instancias son los rótulos **días** y **tiempo**, ambos relacionados con tiempos y sólo el nombre los diferencia. Los dos últimos grupos **MMS** cantidad de MMS, cantidad de MMS dentro de la red y **fecha** de los cuales se identificaron 2 atributos fecha de última recarga, fecha hotline.

Respecto de los atributos asociados a SMS , en base a la información de la ENACOM para la cantidad de SMS enviados desde 2013 hasta el primer trimestre de 2021 en Argentina

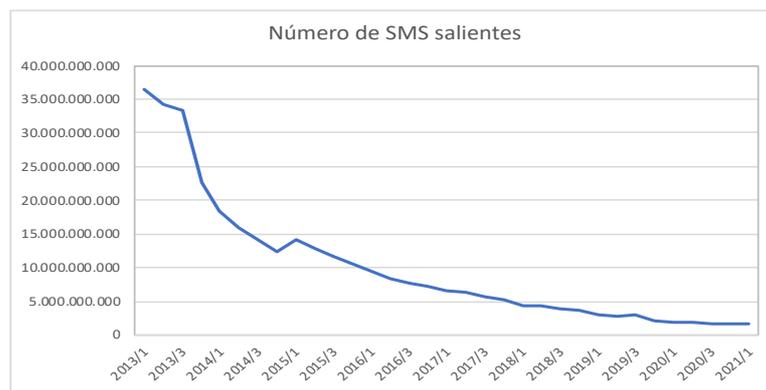


Gráfico 19. Cantidad de SMS en los últimos años³

En el Gráfico 19 se nota un claro descenso del uso por lo que a pesar de contar con 50 instancias en los conjunto de datos en la actualidad podría no ser relevante la elección como lo fue en el pasado.

Capítulo 4

En este capítulo se presentan las conclusiones del trabajo y pruebas de concepto para futuras líneas de investigación

³ Fuente Enacom <https://datosabiertos.enacom.gov.ar/dataviews/240989/numero-de-sms-enviados/>

4.1 Conclusiones

En esta sección se encuentra el resumen de los resultados del trabajo y las posibles futuras líneas de investigación

4.1.1 Resumen de los resultados del trabajo

Existe una gran dispersión entre los resultados de los distintos artículos.

En primer lugar no hay métodos que destaquen sobre otros para la predicción de churn

Las formas de evaluación los resultados no tiene métodos comunes a todos los artículos, si existen procedimientos que son más utilizados y si sucede que los cálculos de los valores se basan en la matriz de confusión.

La cantidad de atributos en los dataset varía de la decena de atributos a los quince mil, de la misma forma hay atributos que se repiten en varios dataset, pero los declarados no hay alguno identificado que se incluya en todos.

En referencia también a las cantidades las muestras que se utilizaron en los distintos artículos varían desde 50 instancias a las 5 millones.

Los artículos si bien se aplican sobre muestras y datos de operadoras reales, en algunos casos están preparados para el análisis, como en algunos ejemplos de dataset la distribución de churn es del 50% , no sería viable una operadora que perdiera la mitad de los clientes, estos ajustes los realizan para que los métodos puedan detectar los patrones de los casos de churn y no los descarten.

Gran proporción de los artículos presentan métodos teóricos para la detección del churn, aplicados sobre dataset de repositorios, pocos casos describen aplicación de procesos reales.

Uno de los artículos presenta consideraciones sobre el peso del cómputo que requiere el procesamiento de la información.

En seis de los artículos hacen referencias expresas que los datos que utilizan de los dataset están compuesto por series temporales.

4.1.2 Futuras líneas de investigación

Del análisis y del resumen del trabajo se puede definir algunos puntos los que se puede desarrollar.

- Utilización de modelos con variación de los atributos de datos incorporando información del mes actual y anteriores.

Son pocos los casos que han utilizado series temporales en las fuentes de datos de la investigación, en este punto deberá existir un compromiso entre un período definido que determine un patrón de posible abandono y un período mayor que oculte este comportamiento. Esta forma de preparación del dataset podría predecir mejor un comportamiento que la de una muestra temporal.

- Identificar atributos influyentes en la detección del churn.

De los métodos que se utilicen deberá producir como salida no sólo al detección del posible churn sino también proporcionar causas que afectaron las decisiones para que las organizaciones puedan actuar en forma preventiva para reducir el churn. Métodos como árboles de decisión presentan los atributos en que se realizan las bifurcaciones en la que dividen los subconjuntos de datos que permitirán interpretar cuales son las razones por la que los clientes toman la decisión de abandono. Para otros métodos será necesario el post-procesamiento para poder detectar estos patrones que derivan en el abandono del servicio. Relacionado con este punto los modelos trabajan sobre valores numéricos. La naturaleza de algunos atributos lo son, costo de servicios, cantidad de llamadas ,etc., pero otros atributos corresponden a clasificaciones que para poder operar sobre ellos será necesario convertir los valores categóricos en numerales y en este sentido el tratamiento de estos atributos deberá obedecer a otras reglas para definir el peso que tome en la predicción de la decisión.

- Verificar la validez del modelo a lo largo del tiempo.

La posible acción de retención podrá modificar el comportamiento (una alteración deseada), y también podría existir algún tipo de ciclo a lo largo del tiempo por lo que sería necesario testear el modelo si es necesario definirlo para cada período o permitiría reutilizarlo n veces siendo necesario determinar dicho n.

4.2 Pruebas de concepto

El principal insumo para el desarrollo del trabajo será contar con los datos necesarios para los distintos ensayos que se deberán realizar

A diferencia de muchos de los casos de los distintos artículos, la fuente de datos no está preparada para realizar las pruebas necesarias.

De la información a la que se tiene acceso, la misma está segmentada por categorías. Uno de los segmentos de información está asociada a los contactos que han tenido los clientes

con la organización, en un caso teniendo como referencia los tiempos que duraron los contactos y por otro lado la cantidad de estos contactos.

Uno de los primeros puntos a tener en cuenta son la cantidad de registros que se tienen por cada uno de estos segmentos. Para todos los datos el rango temporal corresponde a 12 meses. Uno de los conjuntos de datos relacionado a la cantidad de interacciones contiene 1.653.310 registros, mientras el que registra las duraciones tiene 2.527.732. esto significa que no todos los contactos se tiene registro de la misma información. Para el caso del estudio, se trabajará con los datos más completos que se obtengan, en el caso de una aplicación del procedimiento en producción se deberá completar la información faltante.

Para un primer trabajo sobre los datos para la unificación se cargaron los datos individuales sobre un motor de bases de datos mysql (versión 8.0.26-arm64) en un equipo personal con 16 Gb de memoria y disco de estado sólido. La primera estrategia que se uso fue la creación de una tabla que estuviera compuesta por los campos de las tablas, basada en una y completando con los campos no comunes, los campos comunes correspondían a la identificación del suscriptor, el segmento de la llamada, el sentido de la llamada, la provincia y el mes. Como siguiente paso se insertaron los registros de una de las tablas (`insert into .. select`) y luego completar los datos con una actualización (`update TablaConjunta, 2da_tabla TablaConjunta.dato=2da_tabla.dato where ...`). Debido a la cantidad de registros, la sentencia se lanzó y estuvo ejecutándose por un periodo extenso, en una de las últimas verificaciones que estuviera corriendo reportó una duración de 808.008 seg. Esto corresponde a más de 9 días. En este punto se abortó la instrucción y se decidió intentar por otra estrategia.

Una de las opciones era trabajar con tablas temporales fraccionando por provincia y meses, lo que llevaba el trabajar con la tabla de 2,5 Millones de registros a segmentos que iban de los 50.000 a los 130.000 registros cada una.

Una segunda alternativa que se probó para la unificación de las tablas fue la de poblar la tabla conjunta pero con la información de ambas tablas simultaneas (`insert into ... select from Tabla1 inner join Tabla2`). Esta operación arrojó como resultado una demora de 12 segundos.

El resto de la información con la que se cuenta corresponde a característica de los terminales a partir de las actualizaciones reportadas de unos 100 mil registros, información de tráfico 11 millones de registros, información de churn de los usuarios 13 millones de registros e información de facturación con 13 millones de registros.

La cantidad de información con la que se va a tener que trabajar implica que va a ser necesario definir diferentes estrategias, registrando los distintos pasos y buscando la que

minimice el tiempo de armado de los conjuntos de datos, pensando en una posible aplicación práctica, los ciclos de preparación de la información y posterior procesamiento deberían ser lo más acotado para que el análisis y las decisiones a implementar puedan llegar a tiempo para que sean efectivas.

Referencias

- Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), 28.
<https://doi.org/10.1186/s40537-019-0191-6>
- Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. (2019). Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry. *arXiv:1901.06091 [cs, stat]*. <http://arxiv.org/abs/1901.06091>
- Ahn, J., Hwang, J., Kim, D., Choi, H., & Kang, S. (2020). A Survey on Churn Analysis in Various Business Domains. *IEEE Access*, 8, 220816-220839.
<https://doi.org/10.1109/ACCESS.2020.3042657>
- Alboukaey, N., Joukhadar, A., & Ghneim, N. (2020). Dynamic behavior based churn prediction in mobile telecom. *Expert Systems with Applications*, 162, 113779.
<https://doi.org/10.1016/j.eswa.2020.113779>
- Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94, 290-301. <https://doi.org/10.1016/j.jbusres.2018.03.003>
- Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 237, 242-254. <https://doi.org/10.1016/j.neucom.2016.12.009>
- Amin, A., Shah, B., Khattak, A. M., Baker, T., Rahman Durani, H. ur, & Anwar, S. (2018). Just-in-time Customer Churn Prediction: With and Without Data Transformation. 2018 IEEE Congress on Evolutionary Computation (CEC), 1-6.
<https://doi.org/10.1109/CEC.2018.8477954>
- Azeem, M., Usman, M., & Fong, A. C. M. (2017). A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems*, 66(4), 603-614.
<https://doi.org/10.1007/s11235-017-0310-7>
- Dalvi, P. K., Khandge, S. K., Deomore, A., Bankar, A., & Kanade, V. A. (2016). Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. 2016 Symposium on Colossal Data Analysis and Networking (CDAN), 1-4.
<https://doi.org/10.1109/CDAN.2016.7570883>
- De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760-772. <https://doi.org/10.1016/j.ejor.2018.02.009>

- Hassonah, M. A., Rodan, A., Al-Tamimi, A., & Alsakran, J. (2019). Churn Prediction: A Comparative Study Using KNN and Decision Trees. 2019 Sixth HCT Information Technology Trends (ITT), 182-186. <https://doi.org/10.1109/ITT48889.2019.9075077>
- Hu, J., Zhuang, Y., Yang, J., Lei, L., Huang, M., Zhu, R., & Dong, S. (2018). pRNN: A Recurrent Neural Network based Approach for Customer Churn Prediction in Telecommunication Sector. 2018 IEEE International Conference on Big Data (Big Data), 4081-4085. <https://doi.org/10.1109/BigData.2018.8622094>
- Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., & Zeng, J. (2015). Telco Churn Prediction with Big Data. Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 607-618. <https://doi.org/10.1145/2723372.2742794>
- Idris, A., Iftikhar, A., & Rehman, Z. ur. (2019). Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. Cluster Computing, 22(3), 7241-7255. <https://doi.org/10.1007/s10586-017-1154-3>
- Ismail, M. R., Awang, M. K., Rahman, M. N. A., & Makhtar, M. (2015). A Multi-Layer Perceptron Approach for Customer Churn Prediction. International Journal of Multimedia and Ubiquitous Engineering, 10(7), 213-222. <https://doi.org/10.14257/ijmue.2015.10.7.22>
- Jain, H., Khunteta, A., & Srivastava, S. (2021). Telecom churn prediction and used techniques, datasets and performance measures: A review. Telecommunication Systems, 76(4), 613-630. <https://doi.org/10.1007/s11235-020-00727-0>
- Khan, F., & Kozat, S. S. (2017). Sequential churn prediction and analysis of cellular network users—A multi-class, multi-label perspective. 2017 25th Signal Processing and Communications Applications Conference (SIU), 1-4. <https://doi.org/10.1109/SIU.2017.7960659>
- Kitchenham, B. A., & Charters, S. (s. f.). Guidelines for performing Systematic Literature Reviews in Software Engineering | BibSonomy. Recuperado 26 de julio de 2021, de <https://www.bibsonomy.org/bibtex/aed0229656ada843d3e3f24e5e5c9eb9>
- Kostić, S. M., Simić, M. I., & Kostić, M. V. (2020). Social Network Analysis and Churn Prediction in Telecommunications Using Graph Theory. Entropy, 22(7), 753. <https://doi.org/10.3390/e22070753>
- Mahajan, V., Misra, R., & Mahajan, R. (2015). Review of Data Mining Techniques for Churn Prediction in Telecom. Journal of Information and Organizational Sciences, 39(2), 183-197.
- Mitrović, S., Baesens, B., Lemahieu, W., & De Weerd, J. (2018). On the operational efficiency of different feature types for telco Churn prediction. European Journal of Operational Research, 267(3), 1141-1155. <https://doi.org/10.1016/j.ejor.2017.12.015>

- Penetración nacional de la telefonía móvil (accesos por cada 100 habitantes) · Datos Abiertos ENACOM. (s. f.). Recuperado 1 de marzo de 2020, de <https://datosabiertos.enacom.gob.ar/dataviews/241158/penetracion-nacional-de-la-telefoniamovil-accesos-por-cada-100-habitantes/>
- Portabilidad numérica móvil: Portaciones netas mensuales por operador · Datos Abiertos ENACOM. (s. f.). Recuperado 26 de septiembre de 2022, de <https://datosabiertos.enacom.gob.ar/visualizations/29884/portabilidad-numerica-movil-portaciones-netas-mensuales-por-operador/>
- Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, 7, 60134-60149. <https://doi.org/10.1109/ACCESS.2019.2914999>
- Umayaparvathi, V., & Iyakutti, K. (s. f.-a). A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. 03(04), 6.
- Umayaparvathi, V., & Iyakutti, K. (s. f.-b). Automated Feature Selection and Churn Prediction using Deep Learning Models. 04(03), 9.
- Umayaparvathi, V., & Iyakutti, K. (s. f.-c). Automated Feature Selection and Churn Prediction using Deep Learning Models (N.o 03). 04(03), 9.
- Umayaparvathi, V., & Iyakutti, K. (2016). Attribute selection and Customer Churn Prediction in telecom industry. 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 84-90. <https://doi.org/10.1109/SAPIENCE.2016.7684171>
- Wang, C., Li, R., Wang, P., & Chen, Z. (2017). Partition cost-sensitive CART based on customer value for Telecom customer churn prediction. 2017 36th Chinese Control Conference (CCC), 5680-5684. <https://doi.org/10.23919/ChiCC.2017.8028259>
- Zhu, B., Baesens, B., & vanden Broucke, S. K. L. M. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, 408, 84-99. <https://doi.org/10.1016/j.ins.2017.04.015>

Apéndice A: Listado de estudios primarios utilizados en el SMS

- Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), 28.
<https://doi.org/10.1186/s40537-019-0191-6>
- Ahmed, A. A. Q., & Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly based classification. *Egyptian Informatics Journal*, 18(3), 215-220.
<https://doi.org/10.1016/j.eij.2017.02.002>
- Ahmed, A., & Linen, D. M. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1-7. <https://doi.org/10.1109/ICACCS.2017.8014605>
- Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. (2019). Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry. *arXiv:1901.06091 [cs, stat]*. <http://arxiv.org/abs/1901.06091>
- Ahn, J., Hwang, J., Kim, D., Choi, H., & Kang, S. (2020). A Survey on Churn Analysis in Various Business Domains. *IEEE Access*, 8, 220816-220839.
<https://doi.org/10.1109/ACCESS.2020.3042657>
- Alboukaey, N., Joukhadar, A., & Ghneim, N. (2020). Dynamic behavior based churn prediction in mobile telecom. *Expert Systems with Applications*, 162, 113779.
<https://doi.org/10.1016/j.eswa.2020.113779>
- Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94, 290-301. <https://doi.org/10.1016/j.jbusres.2018.03.003>
- Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 237, 242-254. <https://doi.org/10.1016/j.neucom.2016.12.009>
- Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., Hawalah, A., & Hussain, A. (2016). Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study. *IEEE Access*, 4, 7940-7957.
<https://doi.org/10.1109/ACCESS.2016.2619719>
- Amin, A., Shah, B., Khattak, A. M., Baker, T., Rahman Durani, H. ur, & Anwar, S. (2018). Just-in-time Customer Churn Prediction: With and Without Data Transformation. *2018 IEEE Congress on Evolutionary Computation (CEC)*, 1-6. <https://doi.org/10.1109/CEC.2018.8477954>
- ApurvaSree, G. M., Ashika, S., Karthi, S., Sathesh, V., Shankar, M., & Pamina, J. (2019). *Churn Prediction in Telecom using Classification Algorithms* (N.º 1). 2(1), 16.
- Awalludin, Adiwijaya, Bijaksana, M. A., Huda, A. F., & Muslim, L. K. (2017). Churn prediction on fixed broadband internet using combined feed-forward neural network and SMOTEBoost

- algorithm. *2017 5th International Conference on Information and Communication Technology (ICoICT)*, 1-6. <https://doi.org/10.1109/ICoICT.2017.8074672>
- Azeem, M., Usman, M., & Fong, A. C. M. (2017). A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems*, 66(4), 603-614. <https://doi.org/10.1007/s11235-017-0310-7>
- Bi, W., Cai, M., Liu, M., & Li, G. (2016). A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn. *IEEE Transactions on Industrial Informatics*, 12(3), 1270-1281. <https://doi.org/10.1109/TII.2016.2547584>
- Brândușoiu, I., Todorean, G., & Beleiu, H. (2016). Methods for churn prediction in the pre-paid mobile telecommunications industry. *2016 International Conference on Communications (COMM)*, 97-100. <https://doi.org/10.1109/ICComm.2016.7528311>
- Dahiya, K., & Bhatia, S. (2015). Customer churn analysis in telecom industry. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 1-6. <https://doi.org/10.1109/ICRITO.2015.7359318>
- Dalvi, P. K., Khandge, S. K., Deomore, A., Bankar, A., & Kanade, V. A. (2016). Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, 1-4. <https://doi.org/10.1109/CDAN.2016.7570883>
- De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760-772. <https://doi.org/10.1016/j.ejor.2018.02.009>
- Hassonah, M. A., Rodan, A., Al-Tamimi, A., & Alsakran, J. (2019). Churn Prediction: A Comparative Study Using KNN and Decision Trees. *2019 Sixth HCT Information Technology Trends (ITT)*, 182-186. <https://doi.org/10.1109/ITT48889.2019.9075077>
- Hu, J., Zhuang, Y., Yang, J., Lei, L., Huang, M., Zhu, R., & Dong, S. (2018). pRNN: A Recurrent Neural Network based Approach for Customer Churn Prediction in Telecommunication Sector. *2018 IEEE International Conference on Big Data (Big Data)*, 4081-4085. <https://doi.org/10.1109/BigData.2018.8622094>
- Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., & Zeng, J. (2015). Telco Churn Prediction with Big Data. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 607-618. <https://doi.org/10.1145/2723372.2742794>
- Idris, A., Iftikhar, A., & Rehman, Z. ur. (2019). Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Cluster Computing*, 22(3), 7241-7255. <https://doi.org/10.1007/s10586-017-1154-3>
- Idris, A., & Khan, A. (2017). Churn Prediction System for Telecom using Filter–Wrapper and Ensemble Classification. *The Computer Journal*, 60(3), 410-430. <https://doi.org/10.1093/comjnl/bxv123>

- Ismail, M. R., Awang, M. K., Rahman, M. N. A., & Makhtar, M. (2015). A Multi-Layer Perceptron Approach for Customer Churn Prediction. *International Journal of Multimedia and Ubiquitous Engineering*, 10(7), 213-222. <https://doi.org/10.14257/ijmue.2015.10.7.22>
- Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. *Procedia Computer Science*, 167, 101-112. <https://doi.org/10.1016/j.procs.2020.03.187>
- Jain, H., Khunteta, A., & Srivastava, S. (2021). Telecom churn prediction and used techniques, datasets and performance measures: A review. *Telecommunication Systems*, 76(4), 613-630. <https://doi.org/10.1007/s11235-020-00727-0>
- Khan, F., & Kozat, S. S. (2017). Sequential churn prediction and analysis of cellular network users—A multi-class, multi-label perspective. *2017 25th Signal Processing and Communications Applications Conference (SIU)*, 1-4. <https://doi.org/10.1109/SIU.2017.7960659>
- Khan, Y., Shafiq, S., Naeem, A., Ahmed, S., Safwan, N., & Hussain, S. (2019). Customers Churn Prediction using Artificial Neural Networks (ANN) in Telecom Industry. *International Journal of Advanced Computer Science and Applications*, 10. <https://doi.org/10.14569/IJACSA.2019.0100918>
- Kostić, S. M., Simić, M. I., & Kostić, M. V. (2020). Social Network Analysis and Churn Prediction in Telecommunications Using Graph Theory. *Entropy*, 22(7), 753. <https://doi.org/10.3390/e22070753>
- Mahajan, V., Misra, R., & Mahajan, R. (2015). Review of Data Mining Techniques for Churn Prediction in Telecom. *Journal of Information and Organizational Sciences*, 39(2), 183-197.
- Mishra, A., & Reddy, U. S. (2017). A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. *2017 International Conference on Inventive Computing and Informatics (ICICI)*, 721-725. <https://doi.org/10.1109/ICICI.2017.8365230>
- Mitrović, S., Baesens, B., Lemahieu, W., & De Weerd, J. (2018). On the operational efficiency of different feature types for telco Churn prediction. *European Journal of Operational Research*, 267(3), 1141-1155. <https://doi.org/10.1016/j.ejor.2017.12.015>
- P, B., & G.S, N. (2017). A Review on Churn Prediction Modeling in Telecom Environment. *2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, 1-5. <https://doi.org/10.1109/CSITSS.2017.8447617>
- Pamina, J., Beschi Raja, J., Sam Peter, S., Soundarya, S., Sathya Bama, S., & Sruthi, M. S. (2020). Inferring Machine Learning Based Parameter Estimation for Telecom Churn Prediction. En S. Smys, J. M. R. S. Tavares, V. E. Balas, & A. M. Iliyasa (Eds.), *Computational Vision and Bio-Inspired Computing* (pp. 257-267). Springer International Publishing. https://doi.org/10.1007/978-3-030-37218-7_30

- Prashanth, R., Deepak, K., & Meher, A. K. (2017). High Accuracy Predictive Modelling for Customer Churn Prediction in Telecom Industry. En P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition* (pp. 391-402). Springer International Publishing.
https://doi.org/10.1007/978-3-319-62416-7_28
- Research of indicator system in customer churn prediction for telecom industry. (2021, febrero 3).
<https://ieeexplore.ieee.org/abstract/document/7581567>
- Toor, A. A., Usman, M., Younas, F., & Fong, A. (2020). A Robust Systematic Approach for Ensuring Optimal Telecom Service Delivery. *IEEE Communications Magazine*, 58(8), 49-53.
<https://doi.org/10.1109/MCOM.001.2000258>
- Tunga, M. A., & Karahoca, A. (2015). Detecting GSM churners by using Euclidean Indexing HDMR. *Applied Soft Computing*, 27, 38-46. <https://doi.org/10.1016/j.asoc.2014.11.001>
- Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, 7, 60134-60149.
<https://doi.org/10.1109/ACCESS.2019.2914999>
- Umayaparvathi, V., & Iyakutti, K. (s. f.-a). *A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics* (N.º 04). 03(04), 6.
- Umayaparvathi, V., & Iyakutti, K. (s. f.-b). *Automated Feature Selection and Churn Prediction using Deep Learning Models* (N.º 03). 04(03), 9.
- Umayaparvathi, V., & Iyakutti, K. (2016). Attribute selection and Customer Churn Prediction in telecom industry. *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 84-90. <https://doi.org/10.1109/SAPIENCE.2016.7684171>
- Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. Ch. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1-9. <https://doi.org/10.1016/j.simpat.2015.03.003>
- Vijaya, J., & Sivasankar, E. (2019). An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. *Cluster Computing*, 22(5), 10757-10768. <https://doi.org/10.1007/s10586-017-1172-1>
- Wang, C., Li, R., Wang, P., & Chen, Z. (2017). Partition cost-sensitive CART based on customer value for Telecom customer churn prediction. *2017 36th Chinese Control Conference (CCC)*, 5680-5684. <https://doi.org/10.23919/ChiCC.2017.8028259>
- Zhao, L., Gao, Q., Dong, X., Dong, A., & Dong, X. (2017). K- local maximum margin feature extraction algorithm for churn prediction in telecom. *Cluster Computing*, 20(2), 1401-1409.
<https://doi.org/10.1007/s10586-017-0843-2>
- Zhu, B., Baesens, B., & vanden Broucke, S. K. L. M. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, 408, 84-99.
<https://doi.org/10.1016/j.ins.2017.04.015>

Anexos

Glosario de métodos y menciones

Sigla/Acrónimos	Descripción
AdaB+DS	AdaboostM1+Decision Stump
AdaBoost	Adaptive Boosting
AdaC2	AdaBoost Cost-sensitive 2
AFS	Axiomatic Fuzzy Set
ALBA	Adaptive Load - Balanced Algorithm
ALH	Adaptive Local Hyperplane
AMIFS	Adaptative MIFS
ANN	Artificial Neural Networks
Ant minner	Ant minner
ASFDA	Adaptive slow feature discriminant analysis
AUS	Advanced Under- Sampling
AYDASYN	Adaptive Synthetic Sampling
BalanceCascade	BalanceCascade
BBN	Bayesian Belief networks
Bg+RT	Bagging + Random Tree
BM	Bagging Models
Boosting	Boosting
Boosting-SVM	Boosting Support Vector Machine
BRF	Balanced Random Forest
C4.5+RBF	DT 4.5 y Radial Basis Function
CA	Covering Algorithm
CARF	Correlation Attribute Filter Technique
CART	Clasification and Regression Trees
CDDM	Churn Drift Detection Method
CHAID	Chi-squared Automatic Interaction Detector
CLUS	Cluster-based UnderSampling Algorithm
CMIF	Conditional MIFS
CMIN	Conditional mutual info maximization
CNN	Convolutional Neural Network
CS-SVM	Cost-Sensitive Support Vector Machine
CS-Tree	Cost-Sensitive Trees based on in Instance-Weighting
DBN	Deep Belief networks
DCES	Decision Centric Ensemble Selection
DMEL	Data Mining by Evolutionary Learning
DNN	Deep Neural Network
DSIR	Couple Input Symetrical relevance
DT	Decision Tree
DTW	Dinamic time wrapping
EA	Exhaustive Algorithm
EasyEnsemble	EasyEnsemble
EDDM	Early Drift Detection Method
EDMA	Evolutionary Data Mining Algorithm.
Euclidean HDMR	Euclidean High Dimensional Model Representation

Sigla/Acrónimos	Descripción
EUS	Evolutionary Under-Sampling
Exhaustive CHAD	Exhaustive Chi-squared Automatic Interaction Detector
FCM	Fuzzy C-means
FE_RF&T	feature extraction method using Random Forest and Transduction
FLMME	Fuzzy Local Maximal Marginal Embedding
FNN	fuzzy KNN algorithm
FRNN	Fuzzy-rough nearest neighbors
Fuzzy	Fuzzy
GA	Genetic Algorithm
GAM	Generalized Additive Models
Gaussian SVDD	Gaussian Support Vector Data Description
GBDT	Gradient Boosted Decision Tree
GBM	Gradient Boosting Model
Graph	Graph
Graph energy distribution	Graph energy distribution
GWI	Gaussian Weighted Integration
H-MK-SVM	Hierarchical Multiple Kernel Support Vector Machine
HDMR	High Dimensional Model Representation
HFFC	Hybrid FireFly Classification
HLM	hierarchical linear model
IBK	Instance Balance K - Nearest Neighbour
IBRF	Improved Balance Random Forest
ICAP	Interaction Capping J
ICOTE	Immune Centroids Over- sampling Technique
IG	Information Gain entropy
J48	Decision Tree (WEKA)
JIT-CCP	Just in Time Customer churn prediction
JMI	Joint mutual information
K-MC	K-Means Clustering
KLMM	K- local maximum margin
KNN	k-Nearest Neighbor
KP-SVM	Kernel- Penalized Support Vector Machine
LC	Linear Classifiers
LDA	Latent Dirichlet Allocation
LDA[2]	Linear Discriminant Analysis
LEM2	Learning from Example Module, version 2
LI-BLINEAR	Library for Large Linear Classification
LIBFM	Library for factorization machines
LIBSVM	Library for Support Vector Machines
LLM	Logit Leaf Model
LMT	Logistic Model Tree
LP	Label Propagation
LR/RA	Logistic Regression /Regression Analysis
LSTM	Long Short-Term Memory
LSTMNN	Long Short Term Memory Neural Network

Sigla/Acrónimos	Descripción
LWL	Locally Weighted Learning
MIFS	Mutual information feature selection
MIM	Mutual information maximization
MINRED	Minimum redundancy
MLN	Markov Logic Network
MLP	Multi Layer Perceptron
MMC	Maximize Margin Criterion
MOGA	Multi-Objective GA
MOPSO	Multi-Objective Particle Swarm Optimiza- tion
MPGA	Multi Population Genetic Algorithm
MRMR	Minimum-Redundancy Maximum Relevancy
MTDF	Mega-Trend Diffusion Function
MWMOTE	Majority Weighted Minority Oversampling Technique
NA	Network analysis
NB	Naive Bayes
Nnge	Non-Nested Generalisation
NSGA	Non-dominated Sorting GA
OOPM	Orientation Ordering Pruning Method
OSS	One-Sided Selection
OverBagging	OverBagging
OWANN	Fuzzy-rough ownership
Page Rank	Page rank
Parzen-density estimation	Parzen-density estimation
PCA	Principal Component Analysis
PLS	Partial Least Squares
pLSTM	(product layer before recurrent layer) Long Short Term Memory
pRNN	product-based Recurrent Neural Network
PROMETHEE II	PROMETHEE II method
PSO	Particle Swarm Optimization
PWI	Probability Weighted Integration
QUEST	Quick, Unbiased, Efficient Statistical Tree.
RBBagging	Roughly Balanced Bagging
RC	Random Committee
RDDM	Reactive Drift Detection Method
RDR	Ripple Down Rules
RELIEF	Conditional infomax feature extraction
RF	Random Forest
RGA	Rule Generation Algorithms
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
ROS	Random Oversampling
Rot F.	Rotation Fores
Rotboost	Rotboost

Sigla/Acrónimos	Descripción
RRC	Ridge Regression Classifier
RST	Rough Set Theory
RT	Random Tree
RUS	Random Undersampling
RUSBoost	RUSBoost
SAHT	Status Diagram w Hipotesys Test
SCM	Subtractive Clustering Method
SDSCM	Semantic Driven Subtractive Clustering Method
Sender Rank	Sender Rank
SF	Similarity Forest
SGB	Stochastic Gradient Boosting
SGDm	Stochastic Gradient Descent Model
SMOTE	Synthetic Minority Oversampling Technique
SMOTE-ENN	SMOTE + Edited Nearest Neighbor Undersampling
SMOTE-Tomek	SMOTE + Tomek Link Undersampling
SMOTEBagging	SMOTEBagging
SMOTEBoost	SMOTEBoost
SOM	Self-Organizing Maps
STEPD	Statistical test of equal proportions
SVM	Support Vector Machine
TAN	Tree Augmented Naive Bayes
TL-DeepE	Transfer Learning and Ensemble-based Meta-Classification
TOPSIS+	Technique for Order of Preference by Similarity to Ideal Solution
TRkNN	Top-N Reverse K-Nearest Neighbor
TSSA	Time Series Survival Analysis
UnderBagging	UnderBagging
VPNN	voted perception neural network
VQNN	Vaguely Quantified Nearest Neighbors
WK-M	Weighted K Means
WRF	Weighted Random Forest
XGBOOST	eXtreme Gradient Boosting

Atributos identificados

Clasificación 2	Clasificación 1	Descripción	Instancias
valor	estático	3G	1
valor	variable	amigos fuera de la red	2
duración	variable	antigüedad en meses de la línea GSM	1
valor	variable	atributos del headset	1
valor	estático	backup on line	1
valor	variable	bandera de página de display	1
valor	variable	bandera programa hghned	1
valor	variable	cambio de método de pagos	1
valor	variable	cambio de números	1
valor	variable	cambio de tarifas	1
valor	variable	canales	2
cantidad	llamadas	cantidad de llamadas durante la hora pico	1
cantidad	SMS	cantidad de otros MO SMS	1
cantidad	llamadas	cantidad de llamadas de vecinos	1
cantidad	SMS	cantidad cargos SMS salientes	1
cantidad	variable	cantidad de GSM propio	1
cantidad	SMS	cantidad de facturas de SMS	1
cantidad	estático	cantidad de hijos	1
cantidad	SMS	cantidad de información on demand de SMS	1
cantidad	variable	cantidad de lista negra	1
cantidad	llamadas	cantidad de llamadas larga distancia	1
cantidad	llamadas	cantidad de llamadas	4
cantidad	llamadas	cantidad de llamadas de roaming	1
cantidad	llamadas	cantidad de llamadas de voz	2
cantidad	llamadas	cantidad de llamadas dentro de la red	1
cantidad	llamadas	cantidad de llamadas durante el fin de semana	1
cantidad	llamadas	cantidad de llamadas en el día	1
cantidad	llamadas	cantidad de llamadas fuera de la red	1
cantidad	llamadas	cantidad de llamadas al 10010	1
cantidad	llamadas	cantidad de llamadas fuera de la hora pico	1
cantidad	llamadas	cantidad de llamadas locales	2
cantidad	MMS	cantidad de MMS	1
cantidad	MMS	cantidad de MMS con China Mobile	1
cantidad	MMS	cantidad de MMS con China Telecom	1
cantidad	MMS	cantidad de MMS dentro de la red	1
cantidad	MMS	cantidad de MO MMS	2
cantidad	MMS	cantidad de MO MMS rouming internacional	1
cantidad	SMS	cantidad de MO SMS	2
cantidad	SMS	cantidad de MO SMS con china Mobil	1
cantidad	SMS	cantidad de MO SMS con china telecom	1
cantidad	SMS	cantidad de MO SMS de rouming internacional	1
cantidad	SMS	cantidad de MO SMS dentro de la red	1
cantidad	MMS	cantidad de MT MMS	2
cantidad	SMS	cantidad de MT SMS	2
cantidad	variable	cantidad de pagos vencidos	1

Clasificación 2	Clasificación 1	Descripción	Instancias
cantidad	variable	cantidad de prestamos	1
cantidad	SMS	cantidad de proveedores de servicio de SMS	1
cantidad	variable	cantidad de recarga	2
cantidad	SMS	cantidad de SMS	4
cantidad	SMS	cantidad de SMS diario	1
cantidad	variable	cantidad de ultima recarga	1
cantidad	variable	cantidad de vecinos	1
cantidad	llamadas	cantidad llamadas nacionales	1
cantidad	SMS	cantidad de SMS internacional	1
cantidad	SMS	cantidad de SMS local	1
costos	llamadas	cargo ce cantidad de llamadas	1
costos	llamadas	cargo de llamadas	1
costos	SMS	cargo de SMS	1
costos	variable	cargo de sub plan	1
costos	mixto	cargo total del día	6
costos	mixto	cargo total nocturno	3
costos	mixto	cargos de la tardes	6
costos	mixto	cargos de noche	5
costos	mixto	cargos en ingresos	1
costos	mixto	cargos hacia otra red	1
costos	mixto	cargos hogar	1
costos	mixto	cargos mensuales	1
costos	mixto	cargos totales	8
costos	mixto	cargos totales internacionales	8
costos	mixto	cargos trabajo	1
valor	variable	celda comúnmente usada	1
valor	variable	check de método de pago electrónico	1
valor	variable	check de método de pago por correo	1
valor	variable	churn	14
valor	estático	ciudad	1
valor	estático	ciudad natal	2
valor	estático	clase de producto	2
valor	estático	clase de servicio	1
valor	variable	Clasificación ponderada del remitente en el gráfico social	1
valor	estático	código de área	7
valor	estático	código de ocupación	2
valor	estático	contrato un año	1
valor	estático	contrato	2
valor	estático	contrato 2 años	1
valor	estático	contrato mensual	1
costos	variable	contribución estándar para mes 1	1
costos	variable	contribución estándar para mes 2	1
costos	variable	contribución estándar para mes 3	1
costos	variable	costo	1
costos	llamadas	costo actual de la llamada	1
costos	datos	costo de datos diario	1
costos	llamadas	costo de llamadas	5
costos	llamadas	costo de llamadas diario	1
costos	llamadas	costo de llamadas nocturnas	1
costos	datos	costo de servicios en el día	1

Clasificación 2	Clasificación 1	Descripción	Instancias
costos	SMS	costo de SMS diario	1
costos	llamadas	costo del llamado por minuto	1
costos	variable	costo del minuto	2
costos	SMS	costo del sms	1
costos	mixto	costo medio diario	1
costos	llamadas	costo total de la llamadas	2
costos	datos	costo total del GPRS	1
costos	mixto	costo total diario	4
costos	mixto	costo total por mes	2
costos	variable	costo ultima carga	1
costos	MMS	costos del MO MMS	1
cantidad	variable	paquete compradas	1
costos	variable	crédito de usuario	1
valor	variable	cuanto se conocen los amigos entre ellos	1
costos	variable	cuenta	2
valor	variable	datos perdidos en precio handset	1
valor	variable	decremento de la cuenta en mes 1	1
valor	variable	decremento de la cuenta en mes 2	1
valor	variable	decremento de la cuenta en mes 3	1
duración	variable	demora en recarga	1
valor	estático	departamento	1
valor	estático	dependencia	1
valor	estático	dependientes	1
valor	llamadas	detalle de llamadas	1
valor	variable	detalle de servicio de atención al cliente	1
valor	variable	detalle demografía de clientes y personal	1
valor	SMS	detalle SMS	1
valor	variable	detalles de pago y facturas	1
cantidad	variable	día de fin contrato	3
cantidad	variable	días activo	1
cantidad	variable	días activos de recarga	1
cantidad	variable	días activos desde el préstamo	1
cantidad	llamadas	días activos desde la última llamada	1
cantidad	variable	días activos desde la última recarga	3
cantidad	días	Días de cuentas por cobrar (AR)	1
cantidad	días	días de la última transacción de salida	1
cantidad	días	días desde el ultimo uso	1
cantidad	días	días desde la última recarga	1
cantidad	días	días inactivos de datos	1
duración	llamadas	días inactivos de llamadas	1
duración	SMS	días inactivos sms	1
duración	llamadas	duración de llamadas entrantes	1
duración	variable	duración	2
duración	variable	duración cuenta	2
duración	variable	duración de la parada	1
duración	llamadas	duración de llamadas a china mobile	1
duración	llamadas	duración de llamadas	2
duración	llamadas	duración de llamadas de día	1
duración	llamadas	duración de llamadas de larga distancia	1
duración	llamadas	duración de llamadas de rouming	1
duración	llamadas	duración de llamadas de salida	1
duración	llamadas	duración de llamadas de salida de la red	1

Clasificación 2	Clasificación 1	Descripción	Instancias
duración	llamadas	duración de llamadas de voz	2
duración	llamadas	duración de llamadas dentro de la red	1
duración	llamadas	duración de llamadas diario	1
duración	llamadas	duración de llamadas en días festivos	1
duración	llamadas	duración de llamadas en hora pico	1
duración	llamadas	duración de llamadas en línea	1
duración	llamadas	duración de llamadas entrantes	1
duración	llamadas	duración de llamadas fuera de línea	1
duración	llamadas	duración de llamadas libres	2
duración	llamadas	duración de llamadas locales	4
duración	llamadas	duración de llamadas locales de salidas	1
duración	llamadas	duración de llamadas locales internas	1
duración	llamadas	duración de llamadas nacionales	1
duración	llamadas	duración de llamadas de roaming	1
duración	llamadas	duración de los llamadas de vos	1
duración	tiempo	duración en meses	1
duración	tiempo	duración efectiva del paquete	1
duración	llamadas	duración total de llamadas	2
valor	estático	Edad	12
valor	estático	educación	1
valor	variable	esquema de bonos por puntos	2
valor	variable	estadísticas uso internet	1
valor	variable	estado (de la línea)	3
valor	estático	estado civil	2
valor	variable	estado suspensión del cliente	1
valor	variable	estados anormal	1
valor	variable	estatus de conexión de TCP	1
costos	variable	factura	1
costos	variable	factura de ajuste de clientes de salida	1
valor	estático	factura en papel	1
valor	estático	factura residencial	1
costos	variable	factura residual	1
valor	estático	factura sin papel	2
costos	variable	facturación impaga	1
costos	variable	facturación mensual	2
costos	variable	facturas de pago	1
valor	variable	falso si porcentaje de cambio se imputa	1
duración	fecha	fecha de ultima recarga	1
duración	fecha	fecha hotline	1
valor	estático	flag Vas GPRS	1
costos	variable	ganancia media mensual	1
costos	SMS	ganancia SMS	1
costos	estático	gastos mensuales	1
valor	estático	genero	10
valor	variable	grado activo	1
valor	variable	grado de atraso	1
valor	variable	grado de flujo de GPRS	1
valor	llamadas	Grado de gasto de llamadas de roaming	1
valor	llamadas	grado de gasto de llamadas de voz	1
valor	variable	grado de gasto del ultimo mes	1
valor	variable	grado de puntos de miembros	1
valor	variable	grado de sensibilidad del gasto	1

Clasificación 2	Clasificación 1	Descripción	Instancias
valor	variable	grado de tiempos de clase	1
valor	estático	ID área de venta	1
valor	estático	id cliente	2
valor	estático	ID de cargo	1
valor	estático	ID de ciudad	1
valor	estático	ID del producto	1
valor	estático	identificador de llamada	1
valor	variable	inactividad	1
costos	variable	incoming inc ingresos	1
costos	variable	ingresos	1
costos	variable	ingresos CRBT	1
costos	variable	ingresos de IDD	1
costos	variable	ingresos dentro de la red	2
costos	variable	ingresos fuera de la red	3
costos	variable	ingresos hacia otra red	1
costos	variable	ingresos mensuales	2
costos	variable	ingresos paquete GPRS	1
costos	SMS	ingresos por SMS	1
costos	variable	ingresos dentro de la red	1
costos	variable	ingresos FX	1
costos	llamadas	INT_CHG_CNT_M1 costos internacional lamamdas M1	1
costos	tiempo	INT_CHG_MOU_M1 - costos internacional minutos de uso mes 1	1
valor	variable	INT_F_CNT_M1	1
cantidad	llamadas	INT_IC_CNT_M1 - cantidad llamadas entrantes internacional mes 1	1
minutos	tiempo	INT_IC_MOU_M1 - cantidad de minutos de uso entrantes mes 1	1
cantidad	llamadas	INT_OG_CNT_M1 - cantidad llamadas salientes internacional Mes 1	1
minutos	tiempo	INT_OG_MOU_M1 - minutos salientes internacionales Mes 1	1
valor	datos	internet 3G sobre 2G	1
duración	variable	largo de relación de clientes	1
cantidad	SMS	libre sms	2
cantidad	días	life time days	1
valor	variable	limite de crédito	1
cantidad	llamadas	llamadas a atención al cliente	4
cantidad	llamadas	llamadas de tardes	3
cantidad	llamadas	llamadas de entrada	1
cantidad	llamadas	llamadas de salida	2
cantidad	llamadas	llamadas de tarde	1
cantidad	llamadas	llamadas dentro de la red	1
cantidad	llamadas	llamadas distintas	1
cantidad	llamadas	llamadas entrantes	1
cantidad	llamadas	llamadas fuera de la red	2
cantidad	llamadas	llamadas IDD	1
cantidad	llamadas	llamadas internacional fuera de banda	1
cantidad	llamadas	llamadas IP y llamadas larga duración	1
cantidad	llamadas	llamadas libres	1
cantidad	llamadas	llamadas MSISDNS	1

Clasificación 2	Clasificación 1	Descripción	Instancias
cantidad	llamadas	llamadas servicio	1
cantidad	llamadas	llamadas servicio competición	1
cantidad	llamadas	llamados diarios	1
valor	variable	marca teléfono	2
valor	variable	máxima similitud cosine con clientes de otras operadoras	1
valor	variable	máxima similitud cosine con clientes de syriatel	1
valor	variable	máxima similitud jaccard con clientes de otra operadora	1
valor	variable	máxima similitud jaccard con clientes de syriatel	1
cantidad	días	máximo días entre sucesivas recargas para el mes 1	1
cantidad	días	máximo días entre sucesivas recargas para el mes 2	1
cantidad	días	máximo días entre sucesivas recargas para el mes 3	1
cantidad	días	máximo días entre sucesivas recargas para por 3 meses	1
cantidad	días	máximos días entre recargas en mes 1	1
cantidad	días	máximos días entre recargas en mes 2	1
cantidad	días	máximos días entre recargas en mes 3	1
cantidad	días	máximos días entre recargas por 3 meses	1
costos	variable	media de gastos de los últimos 3 meses	1
costos	variable	media de recarga total	1
minutos	llamadas	media de uso de minutos	1
cantidad	variable	meses totales	1
valor	variable	método de pago	2
valor	estático	método de pago tarjeta	1
valor	estático	método de pago transferencia	1
minutos	variable	minuto de uso cargados hacia otra red	1
minutos	variable	minuto de uso de salida hacia otra red	1
minutos	variable	minutos	8
minutos	datos	minutos de datos	1
minutos	llamadas	minutos de llamadas activos de días	2
minutos	llamadas	minutos de llamadas de larga distancia	1
minutos	llamadas	minutos de llamadas durante la hora no pico	1
minutos	llamadas	minutos de llamadas locales	2
minutos	llamadas	minutos de llamada durante la hora pico	1
minutos	llamadas	minutos de llamadas de rouming	1
minutos	llamadas	minutos de llamadas de voz	2
minutos	llamadas	minutos de llamadas durante el fin de semana	1
minutos	llamadas	minutos de salida	1
minutos	llamadas	minutos de tarde	5
minutos	llamadas	minutos de uso hacia otra red	1
minutos	llamadas	minutos dentro de la red	2
minutos	llamadas	minutos diarios	1
minutos	llamadas	minutos fuera de la red	1
minutos	llamadas	minutos IDD	1

Clasificación 2	Clasificación 1	Descripción	Instancias
minutos	llamadas	minutos internacionales	2
minutos	llamadas	minutos libres	2
minutos	llamadas	minutos libres dentro de la red	1
minutos	llamadas	minutos medios mensuales en uso	1
minutos	llamadas	minutos mensuales	1
minutos	llamadas	minutos originados en móvil	1
minutos	llamadas	minutos por mes con suscripción	1
minutos	llamadas	minutos por mes sin suscripción	1
minutos	llamadas	minutos terminados en móvil	1
minutos	llamadas	minutos totales usados	1
duración	datos	Monthly commitment for internet	1
costos	variable	monto de factura del ultimo año	1
costos	variable	monto de facturas del mes n	1
costos	variable	monto de pago promedio	1
costos	variable	monto factura estándar	1
costos	variable	monto factura promedio	1
costos	variable	monto total recarga	2
valor	variable	múltiples líneas	1
valor	variable	nivel del clientes VIP o no VIP	1
valor	variable	nivel grupo	1
valor	variable	número de suerte	1
cantidad	variable	número de amigos conectados con el cliente	1
cantidad	variable	número de amigos del cliente con el que se conectan	1
cantidad	variable	número de amigos y amigos de amigos del cliente	1
cantidad	variable	número de caminos cortos entre cualquiera personas ent la red social a cualquier nodo	1
cantidad	variable	número de días	1
cantidad	llamadas	número de días con llamadas de salida en mes 1	1
cantidad	llamadas	número de días con llamadas de salida en mes 2	1
cantidad	llamadas	número de días con llamadas de salida en mes 3	1
cantidad	llamadas	número de días con llamadas entrantes en el mes 1	1
cantidad	llamadas	número de días con llamadas entrantes en el mes 2	1
cantidad	llamadas	número de días con llamadas entrantes en el mes 3	1
cantidad	variable	número de días de contrato	1
cantidad	variable	número de días hasta que el contrato venza	1
cantidad	variable	número de handset	1
cantidad	variable	número de la suerte	1
cantidad	llamadas	número de llamadas previas hecha por el equipo de retención	1
cantidad	llamadas	número de llamadas nocturnas	1
cantidad	llamadas	número de llamadas salientes	1
cantidad	variable	número de mensajes de correo de vos	5

Clasificación 2	Clasificación 1	Descripción	Instancias
cantidad	variable	número de mensajes	1
cantidad	variable	número de meses impagos	1
valor	variable	número de modelo	1
valor	variable	número de MSISDN entrantes	1
valor	variable	número de MSISDN salientes	1
cantidad	llamadas	número de perdida de llamadas	2
cantidad	variable	número de quejas	3
cantidad	llamadas	número de receptor de llamadas	1
valor	estático	número de red	1
cantidad	variable	número de servicios y mensajes cortos	1
cantidad	SMS	número de SMS enviados en el último mes	1
cantidad	SMS	número de SMS enviados en los 4 a 6 meses	1
cantidad	SMS	número de SMS enviados en los últimos 2 meses	1
valor	variable	número de tenencia del equipo	1
medio	llamadas	número medio de llamadas a director asistida	1
medio	llamadas	número medio de llamadas de vos de salida	1
medio	llamadas	número medio de llamadas de vos entrantes	1
medio	llamadas	número medio de llamadas de vos entrantes y salientes	1
medio	llamadas	número medio de llamadas en espera	1
medio	llamadas	número medio de llamadas a atención a clientes	1
cantidad	llamadas	números de llamadas al servicios al cliente	5
cantidad	variable	números de mes activos	3
valor	variable	números fuera de red	1
cantidad	llamadas	OFF_NET_IC_CNT_M1 cantidad de llamadas entrantes desde fuera de la red del mes	1
minutos	llamadas	OFF_NET_IC_MOU_M1 - minutos de uso de llamadas entrantes desde fuera de la red	1
minutos	llamadas	OFF_NET_OG_MOU_M1- minutos de uso de salida de la red del mes	1
cantidad	llamadas	ON_NET_IC_CNT_M1 - cantidad llamas entrada dentro de la red Mes 1	1
minutos	llamadas	ON_NET_IC_MOU_M1 - minuto de uso de llamadas entrada dentro de la red	1
minutos	llamadas	ON_NET_OG_MOU_M1 minutos de uso salientes dentro de la red Mes 1	1
valor	llamadas	originado en móvil	1
valor	variable	Page Rank ponderado en el gráfico social	1
costos	variable	pagos	1
costos	variable	Pagos totales	1
cantidad	datos	paquete de datos comprados	1
cantidad	datos	paquete de datos entrada y salida	1
cantidad	SMS	paquete de SMS	1
cantidad	variable	paquete de voz	1

Clasificación 2	Clasificación 1	Descripción	Instancias
cantidad	variable	paquete minutos de entrada y salida	1
cantidad	SMS	paquete SMS entrada y salida	1
cantidad	datos	paquetes de bajada de streaming	2
cantidad	llamadas	paquetes de minutos de voz comprado	1
cantidad	SMS	paquetes de SMS de voz compradas	1
valor	estático	partner	3
valor	variable	patrones de uso de clientes	1
cantidad	variable	películas por streaming	1
duración	tiempo	permanencia	13
valor	estático	personas mayores	4
valor	variable	peso promedio de Rank Page y Sender Rank en grafico social	1
valor	variable	plan de correo de voz	8
valor	variable	plan del teléfono	1
valor	variable	plan internacional	9
valor	variable	plan intranet	1
valor	SMS	plan SMS	1
valor	variable	plan tarifario	1
valor	variable	porcentaje de transacciones entre operadores	1
minutos	variable	porcentaje de cambio de minutos de uso	1
valor	llamadas	porcentaje de llamadas fuera de la red	1
valor	llamadas	porcentaje de señales de error/ caída de llamadas	1
costos	variable	precio del producto	1
costos	variable	precio handset	1
valor	datos	promedio de acceso de radio 3G2G	1
medio	variable	promedio de cantidad de recargas en los 3 meses	1
medio	variable	promedio de cargo de uso de minutos	1
medio	variable	promedio de cargos recurrentes	1
medio	llamadas	promedio de duración de llamadas en el último mes	2
medio	llamadas	promedio de duración de llamadas en los últimos 2 meses	1
medio	llamadas	promedio de duración de llamadas en los últimos 4 a 6 meses	1
medio	llamadas	promedio de duración en minutos de llamadas en hora pico	1
medio	variable	promedio de ingreso mensuales	1
medio	llamadas	promedio de llamadas de datos y llamadas de voz	1
medio	llamadas	promedio de llamadas de vos fuera del pico	1
medio	llamadas	promedio de llamadas de voz en hora pico	1
medio	llamadas	promedio de llamadas de voz entrada y salida	1
medio	llamadas	promedio de llamadas internacionales	2
medio	llamadas	promedio de llamadas locales	2
medio	llamadas	promedio de llamados de voz	1
medio	llamadas	promedio de minutos	2

Clasificación 2	Clasificación 1	Descripción	Instancias
medio	llamadas	promedio de minutos de llamadas durante la hora no pico	1
medio	llamadas	promedio de minutos de llamadas durante los fines de semana	1
medio	llamadas	promedio de minutos de llamadas de voz	1
minutos	llamadas	promedio de minutos de uso	1
minutos	llamadas	promedio de minutos internacionales	1
medio	llamadas	promedio de minutos para llamadas nacionales	1
medio	mixto	promedio de máximo precio de recarga en 3 meses	1
medio	mixto	promedio de uso día	1
medio	llamadas	promedio llamadas	2
medio	llamadas	promedio llamadas ob	1
valor	variable	promociones opt in	1
valor	variable	protección de dispositivo	2
costos	variable	rango de recarga mensual	1
minutos	variable	rango de uso de minutos	1
valor	variable	razón de descuento	1
valor	variable	razón del plan	1
costos	variable	recarga de la cuenta	1
costos	variable	recarga total por renovación	1
costos	variable	recargas	2
valor	variable	red compartida en mes 1	1
valor	variable	red compartida en mes 2	1
valor	variable	red compartida en mes 3	1
valor	variable	relación de bajo crédito	1
valor	variable	relación de descuentos	1
valor	variable	relación de falla de clientes	1
valor	variable	relación de gastos	1
valor	llamadas	relación de pérdida de llamadas	2
valor	variable	relación de quemado	1
costos	llamadas	REV_INT_M1	1
costos	SMS	REV_SMS_M1	1
valor	variable	segmento CRM	1
valor	variable	segmento de cliente	3
valor	variable	segmento de valor	1
valor	variable	segmento estilo de vida	1
valor	variable	seguridad online	1
valor	variable	seguro de dispositivo	1
valor	variable	servicio de teléfono	2
valor	datos	servicio internet	2
valor	datos	servicio internet DSL	1
valor	datos	servicio internet por fibra óptica	1
valor	variable	servicio telefonía	1
valor	variable	servicios de valor agregado a clientes	1
valor	variable	si el handset es compatible con la web	1
cantidad	SMS	SMS entrada	1
cantidad	SMS	sms enviados	1
cantidad	SMS	SMS terminados en móviles	1
cantidad	SMS	SMS_CHG_OG_COUNT_M1	1
cantidad	SMS	SMS_F_OG_COUNT_M1	1

Clasificación 2	Clasificación 1	Descripción	Instancias
cantidad	SMS	SMS_OFFNET_CHG_OG_CNT_M1	1
cantidad	SMS	SMS_OFFNET_F_OG_CNT_M1	1
cantidad	SMS	SMS_ONNET_CHG_OG_CNT_M1	1
cantidad	SMS	SMS_ONNET_F_OG_CNT_M1	1
valor	variable	soporte técnico	1
valor	variable	status de conexiones TCP	1
valor	variable	std_vas_arc	1
valor	variable	subscriber MSISDN	1
valor	variable	tamaño de archivos de streaming	2
valor	variable	tamaño de pagina	1
valor	llamadas	tarifa de cantidad de llamadas	1
valor	SMS	tarifa de SMS total	1
valor	variable	tarifa del plan de grupo	1
valor	variable	tarifa del plan	4
valor	SMS	tarifa SMS internacional	1
valor	llamadas	tarifas de llamadas	1
valor	variable	tarifas nuevas	1
cantidad	variable	TCP return time	1
valor	variable	teléfono	2
duración	variable	tiempo de alerta	2
duración	variable	tiempo de parada	1
duración	variable	tiempo de retorno de TCP	1
medio	días	tiempo medio en días entre recargas sucesivas para el mes 1	1
medio	días	tiempo medio en días entre recargas sucesivas para el mes 2	1
medio	días	tiempo medio en días entre recargas sucesivas para el mes 3	1
medio	días	tiempo medio en días entre recargas sucesivas por 3 meses	1
valor	variable	tipo de credencial	1
valor	variable	tipo de marca	1
valor	variable	tipo de producto	1
valor	variable	tipo de quejas	1
valor	variable	tipo de roaming	1
valor	variable	tipo de servicio de call center	1
valor	variable	tipo dispositivo	1
costos	variable	Top-up amount in a day	1
cantidad	variable	total de ingresos de entrada	1
cantidad	variable	total de ingresos de salidas	1
cantidad	llamadas	total de llamadas	3
cantidad	llamadas	total de llamadas de día	4
cantidad	llamadas	total de llamadas entrantes	1
cantidad	llamadas	total de llamadas internacionales	7
cantidad	llamadas	total de llamadas de noche	5
minutos	llamadas	total de minutos	5
minutos	llamadas	total de minutos de día	5
minutos	llamadas	total de minutos de noche	7
minutos	llamadas	total de minutos de salida en mes 1	1
minutos	llamadas	total de minutos de salida en mes 2	1
minutos	llamadas	total de minutos de salida en mes 3	1
minutos	llamadas	total de minutos de salidas	1

Clasificación 2	Clasificación 1	Descripción	Instancias
minutos	llamadas	total de minutos usados en los últimos minutos	1
cantidad	llamadas	total desvinculación de llamadas internacionales	1
cantidad	variable	total desvinculación pasada	1
costos	variable	total ingresos	2
costos	llamadas	total ingresos de llamadas	1
costos	llamadas	total ingresos por voz	1
minutos	llamadas	total minutos de entrada	2
cantidad	SMS	total sms	2
cantidad	variable	total uso de día	1
cantidad	variable	total vas rev	1
valor	variable	TV por streaming	1
valor	estático	ubicación	3
costos	variable	ultimo valor de recarga	2
valor	variable	user effect calculated with pagerank	1
valor	datos	uso de datos	1
cantidad	llamadas	uso de llamadas de ayuda	1
valor	variable	uso servicio de voz	1
valor	datos	uso servicio internet	1
valor	variable	usuarios residentes de china	1
costos	variable	valor de recarga	4
valor	variable	valor del crédito	3
valor	variable	valuación de crédito del cliente	1
costos	variable	variación de gastos	1
costos	variable	variación del gasto diario	1
valor	datos	velocidad de descarga	2
valor	datos	velocidad de subida	2
valor	datos	velocidad internet	1
cantidad	datos	volumen de datos diario	1

Datos Gráfico 16. Métodos de predicción mencionados

Método	Cant.	Método	Cant.	Método	Cant.
LP	1	EDDM	1	LIBFM	2
PLS	1	STEPD	1	EXHAUSTIVE ALGORITMO EA	2
EASYENSEMBLE	1	GRAPH	1	COVERING ALGORITMO CA	2
BALANCECASCADE	1	GRAPH ENERGY DISTRIBUTION	1	LEM2 LA	2
CS-TREE	1	DCES	1	SGB	2
CS-SVM	1	DTW	1	GAM	2
OVERBAGGING	1	SF	1	CNN	2
RBBAGGING	1	TSSA	1	RT	2
SMOTEBAGGING	1	SAHT	1	LLM	2
UNDERBAGGING	1	TAN	1	CART	2
ADABOOST	1	SDSCM	1	QUEST	2
ADAC2	1	SCM	1	CHAID	2
RUSBOOST	1	AFS	1	LIBSVM	2
SMOTEBOOST	1	EUCLIDEAN HDMR	1	MLN	2
BRF	1	CLASICAL HDMR	1	FCM	2
SENDER RANK	1	NNGE	1	DNN	2
HFFC	1	RIDOR	1	MIFS	2
KNN-LR	1	RC	1	PAGE RANK	2
C4.5-RBF	1	TL-DEEPE	1	XGBOOST	3
ADAB+DS	1	DBN	1	DMEL	3
BG+RT	1	KLMM	1	ROT F.	3
MLP	1	MIM	1	ANT MINNER	3
IBK	1	JMI	1	ALBA	3
CARF	1	AMIFS	1	WK-M	3
IG	1	CMIN	1	IBRF	3
LWL	1	ALH	1	BBN	3
LMT	1	ICAP	1	LSTM	3
LC	1	RELIEF	1	BOOSTING	3
SGDM	1	DSIR	1	SOM	4
EXHAUSTIVE CHAD	1	MINRED	1	RST	4
RIPPER	1	CMIF	1	WRF	4
H-MK-SVM	1	RRC	1	GBDT	5
RGA	1	KP-SVM	1	ROTBOOST	5
GAUSSIAN SVDD	1	BOOSTING-SVM	1	PSO	6
PARZEN-DENSITY ESTIMATION	1	FLMME	1	MRMR	6
VQNN	1	JIT-CCP	1	BM	7
FNN	1	MMC	1	GBM	7
FRNN	1	VPNN	1	K-MC	9
FUZZY	1	PRNN	1	GA	15
OWANN	1	PLSTM	1	KNN	15
FE_RF&T	1	MULTI-OBJECTIVE GA	1	NB	21
HLM	1	MOPSO	1	RF	25
PWI	1	TOPSIS+	1	SVM	30
GWJ	1	PROMETHEE II	1	RA o LR	32
CDDM	1	MPGA	1	DT	32
RDDM	1	LI-BLINEAR	2	ANN	34

Gráfico 17. Métodos utilizados en los artículos

Método	Cantidad	Método	Cantidad	Método	Cantidad
DT-C5.0 /4.5/J48	13	hybrid firefly	1	SMOTEBoost + ANN	1
LR	13	RT	1	ANN + ANN	1
RF	12	DS	1	Bagging	1
SVM-RBF /Poly/	10	Adaboost +DS	1	boosting	1
NB	6	Bagging + RT	1	FW-ECP	1
PSO	5	IBK	1	CDDM	1
MLP	5	LWL	1	SDSCM	1
LSTM	4	LLM	1	Logit Boost	1
BPN	4	LMT	1	EI-HDMR	1
KNN	3	DMEL	1	TLDeepE	1
GA	2	SGD	1	DNN	1
GBM	2	GP-adaboost	1	DBN	1
CNN	2	FuzzyNN	1	RRC	1
BPN AdaBoost.M1	1	VQNN	1	S/LFNN	1
AdaBoost.M1	1	FuzzyRoughNN	1	JIT-CCP	1
RST	1	OWANN	1	CART	1
EA	1	AdaBoost+J48	1	MRA	1
COV	1	RFC	1	pRNN	1
LEM2	1	GBC	1	LSTMNN	1
XGBoost	1	VRN	1	pLSTM	1
firefly	1			PNN	1