

**Fundação Getulio Vargas
Escola de Matemática Aplicada**

Pedro Henrique Schneider

**Análise preditiva de Churn com ênfase em
técnicas de Machine Learning: Uma Revisão**

Rio de Janeiro
2016

bib.png

Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Schneider, Pedro Henrique

Análise preditiva de Churn com ênfase em técnicas de Machine learning: uma revisão / Pedro Henrique Schneider. - 2016.
82 f.

Dissertação (mestrado) – Fundação Getúlio Vargas, Escola de Matemática Aplicada.

Orientador: Renato Rocha Souza.

Coorientador: Carlos Saraiva Branco.

Inclui bibliografia.

1. Mineração de dados (Computação). 2. Aprendizado do computador. I. Souza, Renato Rocha. II. Branco, Carlos Saraiva. III. Fundação Getúlio Vargas. Escola de Matemática Aplicada. IV. Título.

CDD – 006.3

Pedro Henrique Schneider

**Análise preditiva de Churn com ênfase em
técnicas de Machine Learning: Uma Revisão**

Dissertação submetida à Escola de Matemática Aplicada como requisito parcial para a obtenção do grau de Mestre em Modelagem Matemática da Informação.

Área de Concentração: Modelagem e Análise da Informação

Orientador: Renato Rocha Souza

Co-Orientador: Antonio Carlos Saraiva Branco

Rio de Janeiro
2016

Agradecimentos

Gostaria de agradecer à minha família, e em especial à minha esposa Mariana, pelo apoio e incentivo ao longo de todos os momentos do mestrado. E ao meu filho Thomas, nascido ao longo do curso de mestrado no dia 28 de outubro de 2014, que trouxe muito trabalho e responsabilidade com seu nascimento, mas, principalmente, muita alegria à família.

Agradeço aos professores e funcionários da Escola de Matemática Aplicada da Fundação Getúlio Vargas, que sempre estiveram dispostos a ajudar e preparados para passar suas experiências e conhecimento para todos os alunos. Agradeço também aos amigos que fizeram parte da minha turma do mestrado. Nossos períodos de convívio e estudo nas salas da Emap serão sempre lembrados.

Agradeço aos meus orientadores Renato e Branco pela contribuição, dedicação e tempo disponibilizado para a elaboração deste trabalho.

ass.png



PEDRO HENRIQUE SCHNEIDER

**ANÁLISE PREDITIVA DE CHURN COM ÊNFASE EM TÉCNICAS DE MACHINE
LEARNING: UMA REVISÃO.**

Dissertação apresentada ao Curso de Mestrado em Modelagem Matemática da Informação da Escola de Matemática Aplicada da Fundação Getúlio Vargas para obtenção do grau de Mestre em Modelagem Matemática da Informação.

Data da defesa: 27/07/2016.

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA


Renato Rocha Souza
Orientador (a)
Antônio Carlos Saraiva Branco
Moacyr Alvim Horta Barbosa da Silva
Alexandre Gonçalves Evsukoff

Resumo

Nas últimas duas décadas, o crescimento da internet e suas tecnologias associadas, vêm transformando a forma de relacionamento entre as empresas e seus clientes. Em geral, a aquisição de um novo cliente custa muito mais caro para uma empresa que a retenção do mesmo. Desta forma, estudos de retenção de clientes, ou gerenciamento do Churn, se tornaram mais importantes para as empresas.

O presente trabalho consiste na revisão e classificação da literatura sobre aplicações de técnicas com ênfase em Machine Learning para construir modelos preditivos de perda de clientes, também chamada de Churn.

O objetivo do trabalho foi reunir o maior número possível de documentos sobre o assunto, dentro da metodologia proposta, e classificá-los quanto às áreas de aplicação, ano de publicação, técnicas de Machine Learning aplicadas, periódicos e repositórios utilizados, nível de influência dos documentos e desta forma trazer à luz os estudos já existentes nesse campo de atuação, consolidando o que há do estado da arte em pesquisas desta área, e de forma significativa contribuir como uma referência para futuras aplicações e pesquisas nesta área.

Embora o trabalho não tenha sido o primeiro na literatura de Machine Learning relacionado a perda ou retenção de clientes na linha de revisão literária, foi o primeiro encontrado com foco em documentos que estudam, não exclusivamente, a perda ou retenção de clientes por técnicas de Machine Learning e sem nenhum tipo de restrições. Da mesma forma foi o primeiro a classificar os documentos por influência através das citações entre os documentos.

Assim, como base final para o trabalho, analisou-se 80 documentos, onde foram encontradas como principais áreas de aplicação: Telecomunicações, Financeiras, Jornais, Varejo entre outras. Constataram-se como técnicas de Machine Learning mais utilizadas para o problema em questão: Regressão Logística, Árvores de Decisão e Redes Neurais, entre outras. E ainda, de acordo com os resultados obtidos, notou-se que ano 2000 tende a ser um marco para esta pesquisa, pois foi a data mais antiga para a qual foi encontrado um artigo nesse trabalho.

Palavras-Chave: *Churn, Análise Preditiva de Churn, Retenção de Clientes, Machine Learning, Aprendizagem de Máquina, Data Mining, Mineração de Dados, Revisão.*

Abstract

In the last two decades, the growth of the Internet and its associated technologies, are transforming the way of the relationship between companies and their clients. In general, the acquisition of a new customer is much more expensive for a company than the retention of a current one. Thus, customer retention studies or Churn management has become more important for companies.

This study represents the review and classification of literature on applications of Machine Learning techniques to build predictive models of customers loss, also called Churn.

The objective of this study was collecting the largest possible number of documents on the subject within the proposed methodology and classifies them as per application areas, year of publication, Machine Learning techniques applied, journals and repositories used and influence level of the documents. And thus, bringing to the light the existing studies in this field of activity, consolidating what is the state of the art of research in this area, and significantly contribute as a reference for future applications and researches in this area.

Although, the study has not been the first in the literature of Machine Learning related to the loss of customer or customer retention in the way of literature review, it was the first, among the ones we have found, with focus on documents studying, not exclusively, loss or retention of customers by Machine Learning techniques, and without any kind of restriction. Furthermore it was the first to classify documents by influence, through the quotations from each document.

As a final database was collected and analyzed 80 documents, from which were found as main application areas: Telecommunications, Financial, Newspapers, Retail, among others. As per Machine Learning techniques applied, the most applied techniques founded related to the problem, were the following: Logistic Regression, Decision Tree and Neural Networks, among others. And based on the results, this kind of study is dated since 2000.

Sumário

1	Introdução	9
2	Referencial Teórico	11
2.1	CRM	11
2.2	Churn	12
2.3	Data Mining	12
2.4	Machine Learning	13
2.4.1	Árvores de Decisão	14
2.4.2	Regressão Logística	18
2.4.3	Redes Neurais	19
2.4.4	Support Vector Machine	23
2.4.5	Random Forest	26
2.4.6	Outras Técnicas Encontradas	27
3	Metodologia de Pesquisa	32
4	Apresentação dos Resultados	43
4.1	Classificação por Técnicas de Machine Learning Empregadas	43
4.2	Classificação por Área de aplicação	45
4.3	Classificação por Ano de Publicação	46
4.4	Classificação por Periódico	47
4.5	Classificação por Repositório	50
4.6	Classificação por Ano & Técnicas de Machine Learning.	51
4.7	Classificação por Área de Aplicação & Técnicas de Machine Learning.	55
4.8	Classificação por Citações	57
5	Conclusões da Pesquisa, Limitações e Trabalhos Futuros.	72
	Referências	75
	Apêndice	84

1 Introdução

Nas últimas duas décadas a grande maioria das empresas em diversas áreas do mercado tem dado significativa importância às áreas de relacionamento com clientes. E não por menos, os estudos e pesquisas nestas áreas têm avançado fortemente. Além disso, o rápido crescimento da internet e de suas tecnologias associadas tem aumentado as oportunidades de ações de marketing das empresas e transformado a forma de relacionamento entre as empresas e seus clientes.

Uma destas áreas trata da retenção de clientes, onde entre outros estudos, encontram-se pesquisas também conhecidas como estudo de Churn de clientes. Predição de Churn ou gerenciamento de Churn, vem há alguns anos preocupando diversas empresas, principalmente aquelas que apresentam um cenário competitivo muito intenso, uma vez que isso representa diretamente perda de receita para as mesmas e tendo em vista que o valor de aquisição de um novo cliente é 5 vezes maior que o custo de retenção do mesmo, segundo mencionado em [Hung, Yen and Wang \(2006\)](#). Também conforme estudo mencionado em [Huang, Kechadi and Buckley \(2012\)](#) a aquisição de um novo cliente pode custar não só 5 vezes mais, mas chegar a 10 vezes mais que a retenção do mesmo. Alguns estudos ainda, como apontado em [Liou \(2009\)](#), citam até mesmo 20 vezes mais como esta diferença de custo. Embora sem consenso de valores, mas sob o consenso de que os custos de aquisição realmente são bem mais altos que os de retenção de clientes, se torna clara a importância dos estudos desta área.

De uma forma mais ampla, a gestão de relacionamento com cliente, CRM da sigla em inglês, consiste em um conjunto de processos de suporte ao negócio, para criar um relacionamento mais longo e mais rentável com os clientes. Informações dos clientes em forma de dados e tecnologia da informação formam a base destes processos, principalmente com o rápido crescimento da internet e tecnologias em geral.

Esta abordagem de CRM visa entender os comportamentos e desejos dos clientes com expectativa que as empresas melhorem seus processos de aquisição, retenção, fidelidade e rentabilidade dos clientes através de informações provenientes dos próprios clientes quando do contato por diversas áreas de relacionamento com as empresas, tais como marketing, vendas, serviços e suportes.

Nesse contexto, Data Mining e Machine Learning representam áreas de conhecimentos com grande abrangência de técnicas para os processos de CRM, e em especial no contexto deste trabalho, para predição do abandono de clientes. E por isso, têm sido utilizados com eficácia na identificação dos clientes mais propensos a deixar a empresa e desta forma promover ações sobre o mesmo aumentando a taxa de sucesso

no processo de retenção.

Esta dissertação tem, portanto, o objetivo de trazer à luz os estudos já existentes nesse campo, consolidando o que há do estado da arte em pesquisas desta área e de forma significativa contribuir como uma referência para futuras aplicações e pesquisas nesta área.

Este trabalho se divide nas seguintes principais seções: 1- Introdução, onde é mencionada a motivação, as etapas e objetivos do trabalho. 2- Referencial Teórico, sessão na qual será abordada mais detalhadamente toda a teoria do assunto, 3- Metodologia de Pesquisa, explicando como e quais critérios foram utilizados ao longo da pesquisa, 4- Apresentação dos Resultados, onde serão expostos os resultados em números, tabelas e gráficos das classificações obtidas com a pesquisa, bem como nossas interpretações dos mesmos, 5- Conclusões da Pesquisa, Limitações e Trabalhos Futuros, é feito um fechamento do trabalho, são expostas algumas limitações sobre o mesmo além de mencionar assuntos em aberto, 6- Referencias, todos os documentos utilizados no trabalho e 7-Apêndices, eventuais planilhas, documentos e informações complementares.

2 Referencial Teórico

Nesta seção pretende-se abordar os principais conceitos teóricos deste trabalho e como os mesmos se relacionam entre si e, desta forma, criar o embasamento teórico para a compreensão da metodologia elaborada e resultados obtidos. Portanto, a seção se inicia com a definição de CRM e o Churn como parte importante deste processo. Segue com um pouco mais de detalhamento do conceito de Churn e em seguida introduz uma interpretação ao conceito de Data Mining, sua relação com Machine Learning, e o motivo pela escolha da ênfase em Machine Learning dada por este trabalho. Feito isso, segue-se por sua vez, com o entendimento obtido quanto à definição de Machine Learning e suas categorias de entradas e saídas de dados. E por fim, são apresentadas as descrições das principais técnicas de Machine Learning encontradas na pesquisa. Nessa ocasião, observa-se que as técnicas já estão ordenadas, antecipando o resultado obtido, quanto a frequência de ocorrência de cada técnica. Ressalta-se também, que para as 5 primeiras técnicas de acordo com este critério, embora ainda que introdutório, suas descrições foram mais detalhadas que as demais e atribuídas subseções individuais para as mesmas.

2.1 CRM

Embora amplamente divulgado e aplicado como uma estratégia de negócio eficaz das empresas, CRM não tem uma definição formal universalmente aceita. [Swift \(2001\)](#) define CRM como “uma abordagem da empresa para entender e influenciar o comportamento do cliente através de comunicações significativas para melhorar a aquisição de clientes, retenção de clientes, lealdade dos clientes e lucratividade dos clientes”. [Kincaid \(2003\)](#) vê CRM como “o uso estratégico da informação, processos, tecnologia, e pessoas para gerenciar o relacionamento do cliente com sua empresa (Marketing, vendas, serviços e suporte) durante todo ciclo de vida do cliente”. [Parvatiyar \(2001\)](#) define CRM como “uma estratégia global de processo de aquisição, retenção, e parceria com os clientes seletos para criar valor superior para a empresa e o cliente. Ela envolve a integração de marketing, vendas, atendimento ao cliente, e as funções da cadeia de suprimentos da organização para alcançar maior eficiência e eficácia na entrega de valor para o cliente”.

Estas são só algumas referências de definições que nos mostram que apesar de não universalmente definido, as definições de CRM estão sempre em concordância, enfatizando a importância do processo desde a aquisição do cliente à retenção do mesmo, ajudando a inteligência do negócio a maximizar o valor daquele cliente para a empresa.

Neste presente estudo, dentro deste processo de CRM, foi dado foco no processo de retenção do cliente, através de estudos de técnicas preditivas de abandono de clientes, também denominado, como mencionado anteriormente, por sua nomenclatura em inglês: Churn.

2.2 Churn

No contexto geral, Churn(taxa de), é a medida do número de indivíduos ou itens que se deslocam para fora de um grupo coletivo ao longo de um período de tempo específico. Quando se refere a uma base de clientes ou assinantes, significa a quantidade de clientes que deixaram a base num determinado período de tempo. Esta medida pode indicar nível de insatisfação, melhores preços praticados por competidores, melhores performances de vendas dos competidores, ou outras razões que tenham a ver com o ciclo de vida do cliente.

Churn está também diretamente ligado ao conceito de tempo de vida do cliente, que aparece em vários estudos de retenção, e conseqüentemente também ao valor agregado ao tempo de vida do cliente (termos em inglês muito utilizados como Customer Life Time e Customer Life Time Value CLTV). Por exemplo, uma taxa de Churn anual de 25%, significa uma taxa média de vida de cliente de 4 anos. Além disso, algumas vezes, estudos de análises de sobrevivência, onde se utilizam termos como “taxa de sobrevivência”, também estão fortemente relacionados com Churn. Neste caso, o mesmo exemplo de uma taxa de Churn de 25%, representa o mesmo que uma taxa de sobrevivência de 75%, isto é, a taxa de sobrevivência costuma significar $1 - a \text{ taxa de Churn}$.

Como o Churn de clientes se caracteriza pelo abandono do cliente aos serviços prestados pela empresa, essa questão tem recebido bastante notoriedade em diversas áreas de negócios como telecomunicações, financeiras, varejos, seguros, jornais e revistas entre outros, pelo fato óbvio em que a perda de cliente representa diretamente a perda de receita das empresas, sem contar a relação de custo já mencionada entre a aquisição e retenção de clientes.

Portanto, diante deste cenário, metodologias e técnicas como as de Data Mining e Machine Learning, se tornaram essenciais na área de Inteligência de Negócios das empresas e vêm sendo utilizadas com o objetivo de diminuir as taxas de Churn.

2.3 Data Mining

Data Mining, é uma grande área da ciência da computação, definido como “o processo de extração de padrões escondidos de grandes conjuntos de dados usando estatística,

matemática, inteligência artificial e técnicas de aprendizado de máquina”, [Berson, Smith and Thearling \(2000\)](#).

[Thearling \(1999\)](#), propôs que Data Mining é a “extração preditiva de informação oculta a partir de grandes bases de dados”.

[Lejeune \(2001\)](#), abordou que as técnicas de mineração de dados “permitem a transformação dos dados brutos em conhecimento do negócio”.

[Institute \(2000\)](#), define a mineração de dados como “o processo de selecionar, explorar e modelar grande quantidade de dados para descobrir padrões de dados previamente desconhecidos para vantagens de negócios”.

Em resumo, vendo estas definições acima, como tantas outras, nota-se que Data Mining tem sido uma ferramenta poderosa nessa missão de extrair e identificar nos dados informações que sejam relevantes e possam trazer vantagens para os negócios.

Hoje em dia já existe uma enorme quantidade de dados disponíveis em cada organização, cuja tendência, ainda assim é de crescimento, o que significa que cada vez mais as técnicas de Data Mining se tornam essenciais para transformar todos esses dados em informações úteis ao negócio.

2.4 Machine Learning

Ainda neste contexto, nesse trabalho como em toda a literatura do assunto, costuma-se alternar entre os termos Machine Learning e Data Mining para expressar grupos de técnicas utilizadas. Isso porque ambas as áreas empregam em geral os mesmos métodos e se sobrepõem de forma significativa. Data Mining e Machine Learning utilizam-se muitas vezes dos mesmos métodos porém com objetivos um pouco diferentes em mente.

Por isso, não se teve a pretensão neste trabalho de tentar definir uma ou outra área, nem tampouco afirmar a diferença entre elas. Contudo, ao longo das pesquisas, assumiram-se algumas interpretações a respeito, como: em Machine Learning, o desempenho é geralmente avaliado com relação à capacidade de reproduzir o conhecimento conhecido, enquanto em Data Mining a tarefa fundamental é a descoberta de conhecimento até então desconhecida. Ou ainda: Machine Learning se concentra na previsão e classificação com base nas propriedades aprendidas a partir de um conjunto de dados de treinamento. Enquanto Data Mining foca-se na exploração dos dados e descobertas de propriedades não conhecidas dos mesmos.

Por estas interpretações, preferiu-se dar mais ênfase ao termo Machine Learning, tendo em vista que o problema em questão se baseia justamente em uma classificação de dados, ou uma previsão dos dados, partindo de dados históricos já conhecidos. Isto é, em outras palavras, através de uma base histórica de clientes que fizeram o Churn,

todos os modelos e métodos, tentam classificar os grupos de clientes existentes de acordo com sua propensão ao Churn e desta forma obter uma previsão e ação sobre estes grupos de maneira diferenciada.

Portanto, assim como já mencionado anteriormente para Data Mining, Machine Learning também pertence a um subgrupo da ciência da computação, que se desenvolveu a partir do estudo de reconhecimento de padrões e de teorias de aprendizagem computacional em inteligência artificial. Machine Learning se caracteriza fortemente pelo uso de modelos e algoritmos de aprendizagem automática, onde os mesmos podem aprender e fazer previsões partindo de dados de entrada conhecidos.

Do ponto de vista das entradas e natureza de aprendizado, as tarefas e problemas relacionados a Machine Learning podem ser classificados em três principais categorias: ***aprendizagem supervisionada***, ***aprendizagem não supervisionada*** e ***aprendizado por reforço***.

Em se tratando do problema de análise preditiva de Churn, são aplicados modelos que são classificados na categoria de **aprendizagem supervisionada**. Nesta categoria o modelo é apresentado a exemplos de entradas onde suas saídas desejadas são conhecidas, ou seja, dados históricos conhecidos fornecidos por fontes externas ao modelo. O objetivo é aprender uma regra geral que mapeia entradas e saídas.

Do outro ponto de vista, isto é, da saída desejada de cada modelo, Machine Learning pode ser categorizado em outros diversos grupos de modelos tais como associação, classificação, clusterização, previsão, regressão, análise de sequências e ainda, uma classe à parte, que seria a de visualização.

Destas categorias, ***classificação***, é um dos modelos mais comuns e mais utilizados de Machine Learning. Em termos básicos, tem o objetivo de identificar a que categoria pertence uma nova observação. Serve para a construção de modelos para prever, por exemplo, comportamentos futuros através da classificação de registros existentes no banco de dados. E Churn, é dos exemplos para o qual estes modelos são adotados.

A seguir, conforme mencionado na introdução desta seção, serão descritas as principais técnicas encontradas na pesquisa.

2.4.1 Árvores de Decisão

Árvores de Decisão é um método habitualmente usado em Machine Learning. O objetivo é o de criar um modelo que prevê o valor de uma variável de destino com base em diversas variáveis de entrada. Uma árvore é uma estrutura do tipo fluxograma em “árvore”, onde cada nó interno (nó não-folha) indica um teste de um atributo, cada ramo representa um resultado do teste, e cada folha nó (ou nó terminal) contém

um rótulo de classe. O nó de nível superior em uma árvore é o nó raiz.

Uma Árvore de Decisão é uma representação simples de método de classificação. Assumimos, como no problema em estudo, que exista apenas uma única variável destino para a classificação. Cada elemento do domínio da classificação é chamado de uma classe, tem-se então uma classe representada por “Churn”, aquelas observações que representam indivíduos que fizeram o Churn e a classe “No Churn”, claramente representando o restante dos indivíduos. Uma Árvore de Decisão ou uma árvore de classificação é uma árvore em que cada um dos nós internos (não folhas) é marcado com uma variável de entrada. As arestas vindas de um nó marcado com uma variável de entrada são rotuladas com cada um dos possíveis valores desta variável. E por último, cada folha de uma árvore é marcada com uma das duas classes mencionadas, “Churn” ou “No Churn”, ou uma distribuição de probabilidade dessas classes.

Uma árvore pode ser “aprendida” dividindo a origem em subconjuntos com base em um teste de valor da variável. Este processo é repetido em cada subconjunto derivado, de uma maneira recursiva chamada partição recursiva. A recursividade é completada quando o subconjunto em um nó tem todos os seus elementos com o mesmo valor da variável de destino, ou quando a divisão não agrega mais valor às previsões. Este processo de indução de cima para baixo da Árvore de Decisão é um exemplo de um algoritmo guloso, e é a estratégia mais comum neste tipo de abordagem.

A figura 1 mostra a representação de uma Árvore de Decisão, para uma classificação de compra, mas que analogamente poderia representar uma situação de Churn, que representa a intenção de compra ou não de computadores em uma loja.

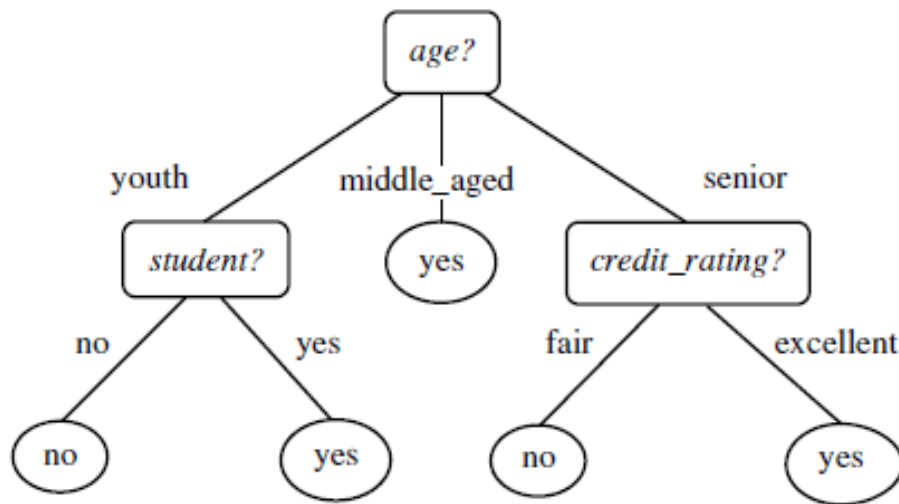


Figura 1: Representação de uma Árvore de Decisão para classificação de compra de computador.

Fonte: [Han and Kamber \(2006\)](#)

Já a figura 2 mostra a saída de uma ferramenta de software analítico, para uma Árvore de Decisão de uma análise preditiva de Churn. Pode-se notar, que na prática, é bem difícil ter folhas (nós terminais) com 100% de uma das classes. Ao invés disso, o que se tem, conforme mencionado, são distribuições de probabilidades das duas classes.

A estratégia de um algoritmo guloso em um método de Árvores de Decisão, em geral, busca escolher em cada passo a variável que melhor divide o conjunto de dados. Como foi dito acima, essa divisão é feita com base em um teste de valor de atributo, e diferentes métricas utilizadas neste teste diferenciam os algoritmos utilizados nas construções das Árvores de Decisão.

Alguns deles e suas métricas são:

- CART: Teste Gini — CART, classificação e árvore de regressão (do inglês, classification and regression tree), utiliza-se da métrica de impureza Gini que é uma medida de quantas vezes um elemento escolhido aleatoriamente a partir do conjunto seria rotulado incorretamente se ele foi marcado aleatoriamente de

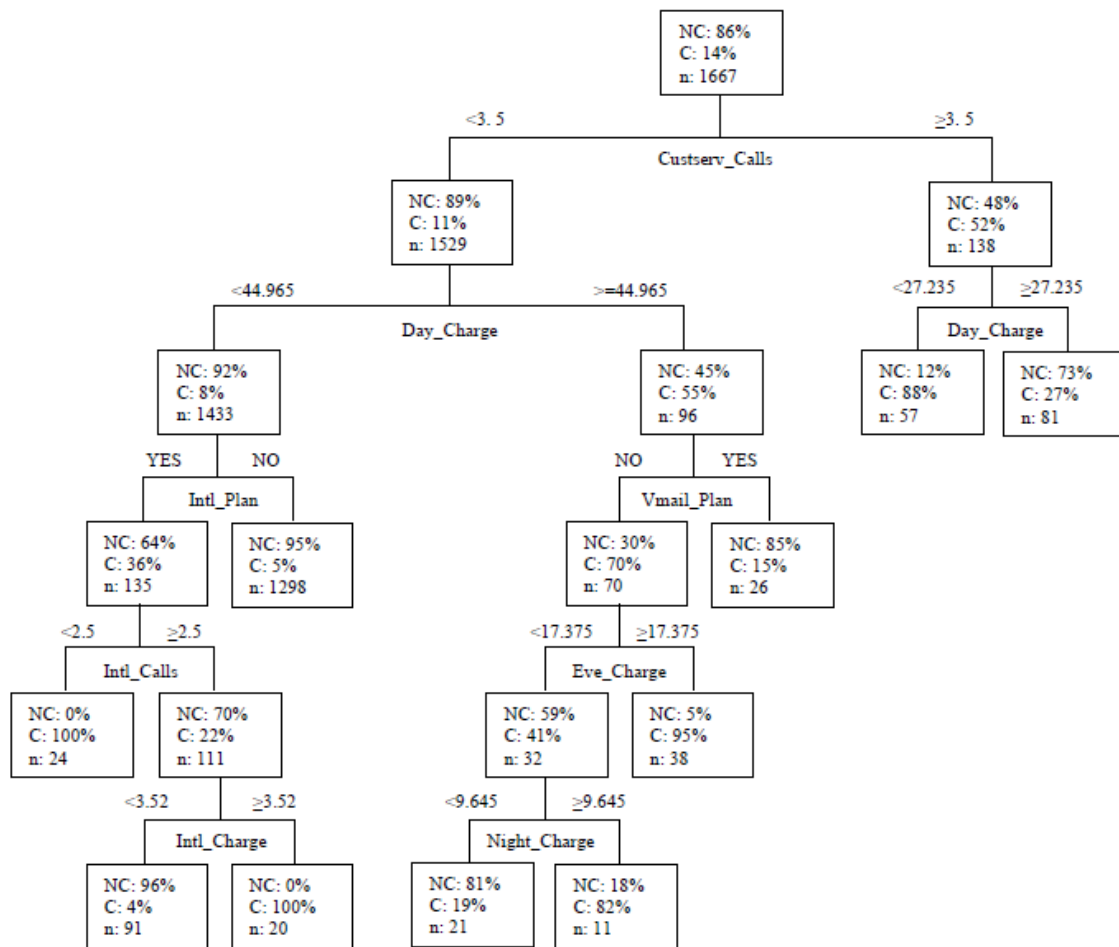


Figura 2: Representação de uma Árvore de Decisão de Análise de Churn em Telecom.
Fonte: [Lima, Mues and Baesens \(2009\)](#)

acordo com a distribuição de rótulos no subconjunto.

- CHAID: Teste de Chi Quadrado — CHAID é um tipo de técnica de Árvores de Decisão, com base em testes de significância ajustado, o teste de Chi-Quadrado. É uma técnica não paramétrica, possui a saída altamente visual e fácil de interpretar. Por outro lado precisa de uma amostra bastante grande para trabalhar de forma eficaz.
- ID3 e C4.5: Teste de Entropia — C4.5 constrói Árvores de Decisão a partir de

um conjunto de dados de treinamento da mesma forma como no ID3, utilizando o conceito de entropia da informação.

Por que os classificadores de Árvores de Decisão são tão populares? A construção de uma Árvore de Decisão em geral não requer muito conhecimento da técnica nem configurações de parâmetros, e, portanto, são apropriados para descoberta de conhecimento exploratória. As Árvores de Decisão podem lidar com alta dimensionalidade dos dados. Sua representação em forma de árvore é intuitiva e, geralmente, fácil de assimilar. As etapas de aprendizagem e classificação de uma Árvore de Decisão são simples e rápidas. Em geral, os classificadores de Árvores de Decisão tem boa precisão. (Han and Kamber, 2006)

2.4.2 Regressão Logística

Regressão logística, ou regressão logit, ou modelo logit, é um modelo de regressão em que a variável dependente é binária ou categórica.

A Regressão Logística mede a relação entre a variável dependente binária e uma ou mais variáveis independentes estimando probabilidades utilizando uma função logística, que é a Distribuição Acumulada Logística. Na interpretação da variável dependente deste método, o erro da Regressão Logística assume uma distribuição logística padrão.

A Regressão Logística é mais um caso, particular, de modelos lineares. A Classificação Linear utiliza o que é, em geral, chamado de limite rígido, isto é, faz a classificação através do sinal de $s = w^T x$,

$$h(x) = \text{sign}(w^T x)$$

Em contra partida a Regressão Linear não utiliza limite algum, ou seja,

$$h(x) = w^T x$$

Já a Regressão Logística, pode ser vista como um caso intermediário à estes dois casos vistos acima. Ou seja, o modelo busca suavemente restringir sua saída para um intervalo de probabilidade entre $[0,1]$. Este modelo é chamado de modelo de Regressão Logística e pode ser representado da seguinte forma:

$$h(x) = \theta(w^T x),$$

Onde θ é chamado de Função Logística,

$$\theta(s) = \frac{e^s}{1+e^s}, \text{ cuja saída está entre 0 e 1.}$$

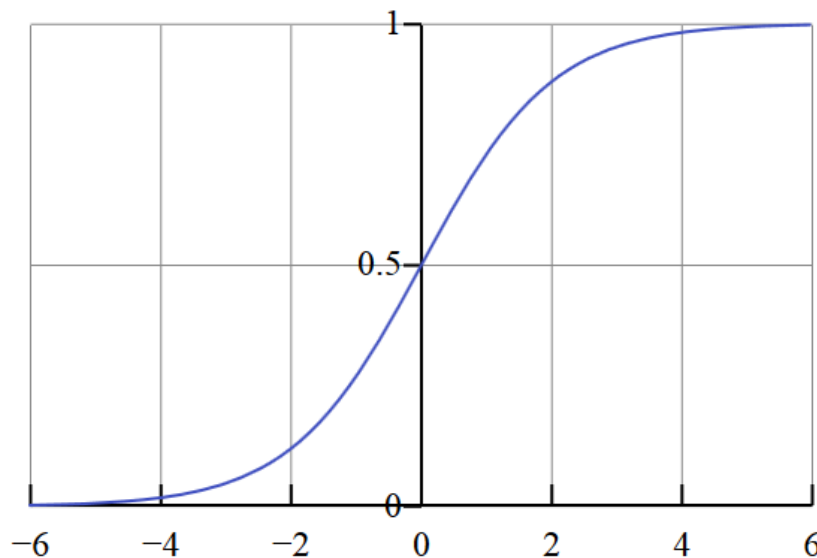


Figura 3: Função Logística no intervalo de s (-6,6).

Fonte: https://en.wikipedia.org/wiki/Logistic_regression#/media/File:Logistic_curve.svg

Um gráfico da função logística no intervalo de s (-6,6) é mostrado na figura 3.

A saída do modelo, como foi dito, pode ser interpretado como a probabilidade para um evento binário. Desta forma, como desejado, o modelo pode ser utilizado para prever no problema de análise de Churn, a probabilidade de “Churn” ou “No Churn”. A Classificação Linear também lida com eventos binários, porém, a diferença é que a “classificação” no modelo logístico pode lidar com a incerteza representada por uma probabilidade com valores intermediários entre 0 e 1, o que na prática, em casos reais, é bastante útil. Aliado a simplicidade do método, a facilidade de interpretação e acurácia satisfatória, faz deste um método bastante utilizado.

2.4.3 Redes Neurais

Em Machine Learning, Redes Neurais Artificiais ou simplesmente Redes Neurais é uma família de modelos inspirados em Redes Neurais biológicas (o sistema nervoso central de animais, em especial o cérebro) e são utilizados para estimar ou aproximar funções que podem depender de um grande número de entradas e são geralmente desconhecidas.

As Redes Neurais são geralmente caracterizadas como sistemas de “neurônios”

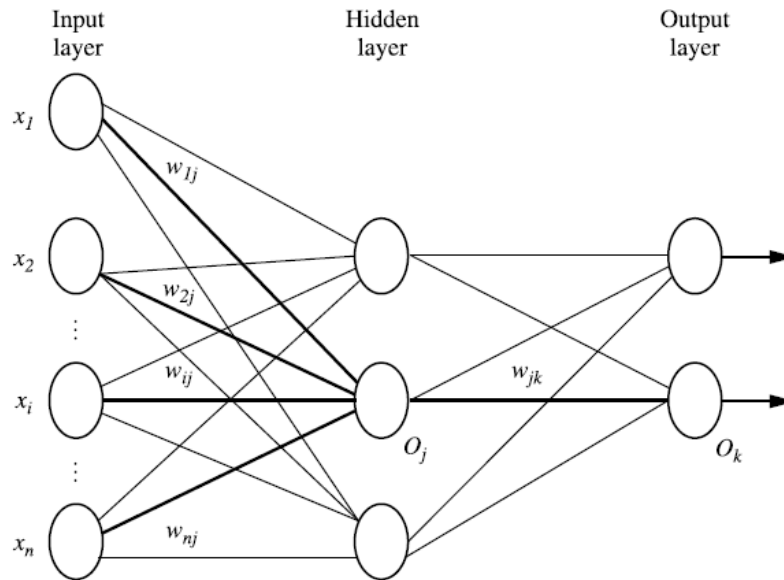


Figura 4: Exemplo de representação de uma Rede Neural de duas camadas.

Fonte: [Han and Kamber \(2006\)](#)

interconectados que trocam mensagens entre si. As conexões têm pesos numéricos que podem ser ajustados com base na experiência, através de um algoritmo de aprendizagem, tornando as Redes Neurais adaptativas às entradas, capazes de aproximar funções não lineares e fundamentalmente capazes de aprender.

Uma rede neural é tipicamente definida por três tipos de atributos:

1. As unidades de processamento, que são os neurônios e os pesos de suas conexões.
2. A estrutura de interligações entre as diferentes camadas de neurônios.
3. O processo ou algoritmo de aprendizagem para atualizar os pesos das interconexões.

A figura 4 apresenta um exemplo de topologia de Redes Neurais de duas camadas, já que a camada de entrada não é contada, pois apenas serve para passar os valores de entrada para a camada seguinte.

No caso dos neurônios, que são os núcleos de unidades processadoras, as principais partes são:

- As sinapses, caracterizadas pelos seus pesos associados;

- A função somadora ou função de combinação;
- A função de ativação;

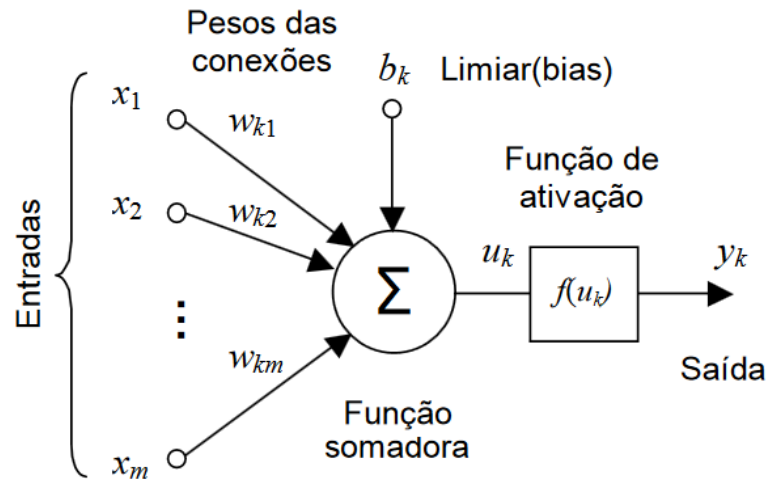


Figura 5: Neurônio.

A figura 5 representa a composição de um neurônio. Nesta representação, o primeiro subscrito k do peso sináptico w_{kj} corresponde ao neurônio pós-sináptico, e o segundo subscrito corresponde à sinapse ligada a ele.

As funções de combinação podem ser de adição simples, lineares ou radiais. As funções lineares, que são mais comumente utilizadas, tratam de uma combinação linear das entradas pelos pesos associados, mais o limiar b_k . Isto é, soma todos os sinais de entrada ponderados pelos pesos das conexões e soma com b_k .

Já as funções de ativação podem ser de três tipos: Degrau, Pseudo-Linear e Sigmoid. E são utilizadas com os propósitos de limitar a saída do neurônio e introduzir não linearidade ao modelo. Dentre estes tipos, as funções do tipo Sigmoid em geral são as mais aplicadas, tendo como exemplos entre elas a função Logística e função Tangente Hiperbólica.

O limiar b_k tem o papel de aumentar ou diminuir a influência do valor de entrada para a ativação do neurônio k .

Matematicamente, resumindo o que foi dito acima, a saída do neurônio k pode ser descrita por:

$$y_k = f(u_k) = f[\sum_{j=1}^m w_{kj}x_j + b_k] \text{ ou } y_k = f(u_k) = f[\sum_{j=0}^m w_{kj}x_j]$$

onde, neste último, x_0 é um sinal de entrada de valor 1 e peso associado $w_{k0} = b_k$.

Do ponto de vista de topologia ou de estrutura de interconexão entre os neurônios, as redes possuem, como foi visto no exemplo acima, uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. O usuário precisa definir, antes da fase de treinamento, o número de unidades em cada camada, assim como o número de camadas escondidas. As redes podem ser classificadas ainda como Feedforward ou Recorrentes. Feedforward quando as redes propagam o sinal em uma única direção, da entrada para a saída. As redes Recorrentes são denominadas desta forma, pois possuem pelo menos um ciclo de realimentação entre os neurônios na rede.

E por último, mas não menos importante, tem-se os paradigmas de aprendizagem das Redes Neurais. A capacidade de aprender associada a uma Rede Neural é uma das mais importantes características destas estruturas. Como foi dito, no contexto de Redes Neurais, a capacidade de aprendizagem ou treinamento, corresponde ao processo de ajuste dos parâmetros livres da rede. Primeiramente, os modelos podem ser subdivididos nos 3 grandes grupos já mencionados: Aprendizagem Supervisionada, Não-Supervisionada e Por Reforço. Em seguida, a distinção se dá pelos algoritmos utilizados para treinamento da rede. Os algoritmos são métodos numéricos responsáveis por encontrar o conjunto de pesos que maximize o ajuste da rede aos dados fornecidos. Como exemplos de algoritmos têm-se: Perceptron (PLA), Gradiente Descendente, Backpropagation, Levenberg-Marquardt, Quasi-Newton entre outros.

As Redes Neurais podem ser utilizadas para problemas de classificação, como o problema de análise preditiva de Churn. Nestes casos, cada unidade de saída pode ser utilizada para representar duas classes. Assim, uma rede construída para o problema de Churn, poderia ter uma única unidade de saída, onde, por exemplo, um valor 1 poderia representar a classe “Churn” enquanto um valor 0 poderia representar a classe “No Churn”.

E de fato, as Redes Neurais têm sido bastante utilizadas para diversas aplicações de classificação, inclusive para a aplicação de Churn. Em geral, elas são criticadas por seus longos períodos de treinamento, pela determinação de parâmetros de forma empírica e, no caso de Churn, principalmente pela falta de interpretabilidade de sua estrutura. Porém, as mesmas têm se mostrado robustas aos ruídos, aos casos onde as relações entre os parâmetros não são bem conhecidas, e apresentam alto grau de paralelismo, acelerando o processamento de grandes bases de dados como normalmente são os casos de Churn. E acima de tudo, têm apresentado resultados com boas precisões.

2.4.4 Support Vector Machine

Em Machine Learning, Support Vector Machine (SVM) são modelos de aprendizagem supervisionados com algoritmos associados que analisam os dados utilizados para fazer classificações lineares e não lineares. Podendo, desta forma, ser utilizados para a classificação do problema de análise preditiva de Churn.

Em poucas palavras, Support Vector Machine (SVM) é um algoritmo que funciona como se segue. Ele usa um mapeamento não linear para transformar os dados de treinamento original em uma dimensão mais elevada. Dentro desta nova dimensão, ele procura o hiperplano de separação ideal linear (isto é, uma “fronteira de decisão” separando as tuplas de uma classe de outra). Com um mapeamento não linear adequado para uma dimensão suficientemente elevada, os dados a partir de duas classes podem sempre ser separados por um hiperplano. O SVM encontra este hiperplano usando os vetores de suporte e suas margens.

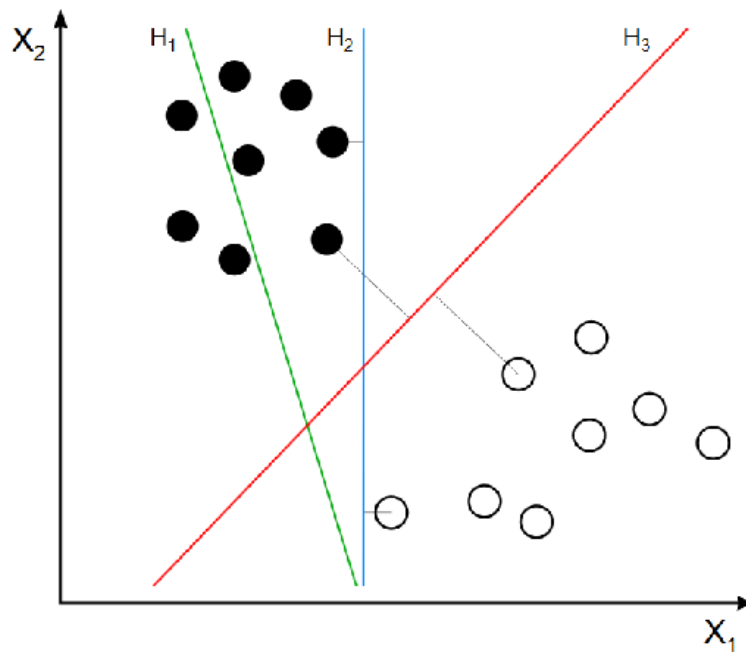


Figura 6: Modelo de SVM e seus Hiperplanos H 's

Fonte: [https://en.wikipedia.org/wiki/Support_vector_machine#/media/File : Svm_separating_hyperplanes%28SVG%29.svg](https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_separating_hyperplanes%28SVG%29.svg)

A figura 6 representa um modelo de Support Vector Machine e alguns hiperplanos de classificação. Como no exemplo da figura, nota-se que o Hiperplano H_1 não separa

de forma correta as duas classes existentes. Já os hiperplanos H_2 e H_3 , fazem essa separação de forma perfeita, porém H_3 a faz com uma “margem de segurança” maior.

Considerando que o problema original pode ser declarado em um espaço de dimensão finita, acontece frequentemente que os conjuntos para discriminação não são linearmente separáveis nesse espaço. Por esse motivo, propõe-se que o espaço de dimensão finita inicial seja mapeado para um espaço de dimensão muito maior, tornando-se, presumidamente a separação mais fácil neste espaço. Neste ponto, entra a utilização das funções Kernel para manter a carga computacional razoável e assegurar que os cálculos possam ser executados mais facilmente em termos das variáveis no espaço original.

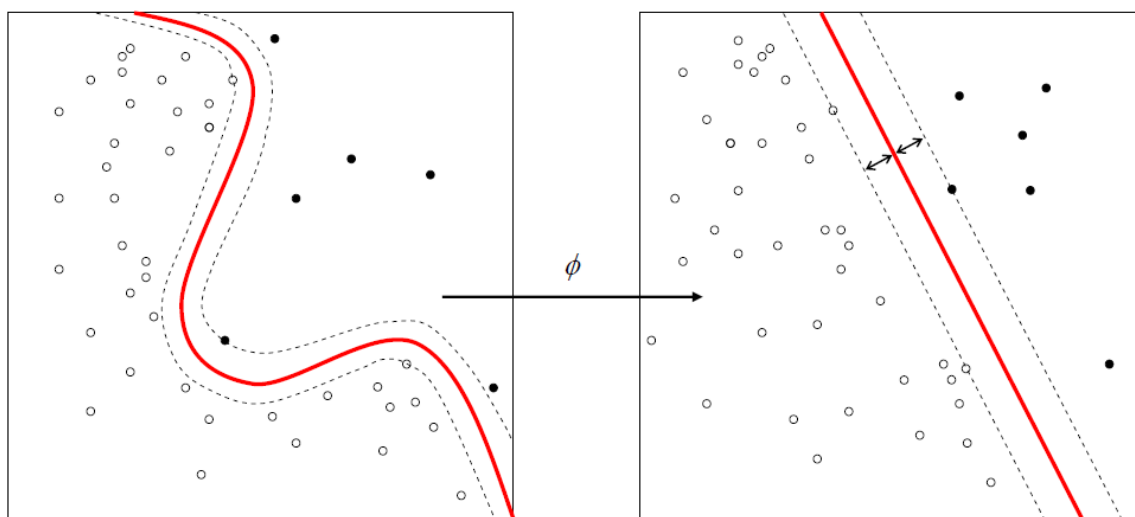


Figura 7: Função Kernel do modelo SVM

Fonte: https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Kernel_Machine.png

A figura 7 representa a função Kernel.

Para um caso linear, nos é dado um conjunto de dados de treinamento de n pontos da forma:

$$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$$

onde os y_i são 1 ou -1 , cada um indicando a classe à qual o ponto de \vec{x}_i pertence. Ou seja, novamente em analogia com o problema em questão, o valor 1 poderia representar a classe “Churn” e o valor -1 a classe “No Churn”. Cada \vec{x}_i é um vetor

de valores reais de dimensão p . Queremos encontrar o hiperplano com a maior “margem” que divide o grupo de pontos \vec{x}_i para os quais $y_i = 1$ do grupo de pontos para os quais $y_i = -1$, a qual é definida de modo que a distância entre o hiperplano e o ponto \vec{x}_i mais próximo de qualquer dos dois grupos seja maximizada.

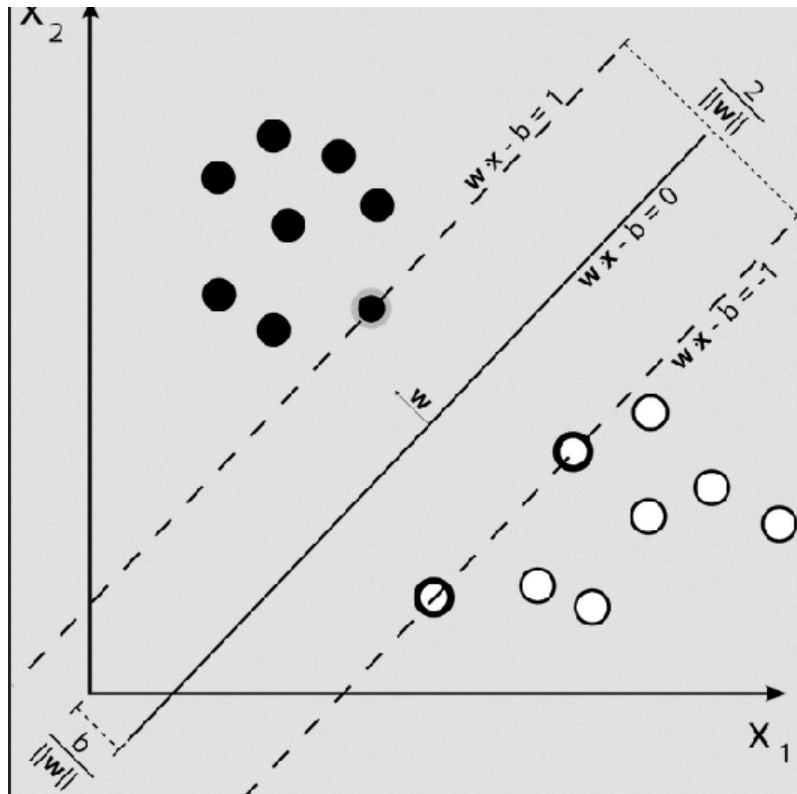


Figura 8: Maximização da margem de um modelo SVM.

Fonte: [https://en.wikipedia.org/wiki/Support_vector_machine#/media/File : Svm_max_separating_plane_with_margin.png](https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_max_separating_plane_with_margin.png)

A figura 8 representa a maximização da distância entre o hiperplano e os pontos mais próximos de cada classe.

Qualquer hiperplano pode ser escrito como o conjunto de pontos \vec{x} satisfazendo:

$$\vec{w} \cdot \vec{x} - b = 0,$$

onde \vec{w} é um vetor normal ao hiperplano (não necessariamente normalizado). O parâmetro $\frac{b}{|\vec{w}|}$ determina o deslocamento do hiperplano desde a origem ao longo do vetor normal.

Uma importante consequência desta descrição geométrica, é que o hiperplano de margem máxima é definido pelos \vec{x}_i mais próximos ao mesmo. E estes \vec{x}_i são então chamados de vetores de suporte (Support Vector).

Os modelos de Support Vector Machine, são mais recentes que os demais vistos até agora, sendo considerado um método promissor de classificação. Embora eles sejam considerados lentos, eles podem lidar, como foi visto, com bases de dados de grandes dimensões (o que é essencialmente importante para estudos de Churn) e têm apresentando resultados altamente precisos. (Han and Kamber, 2006)

2.4.5 Random Forest

Random Forest é um método de aprendizagem combinado, utilizado, entre outras tarefas, para classificação. O método opera através da construção de um grande número de Árvores de Decisão na fase de treinamento e retorna como resultado a classe que é a moda (estatisticamente) dentre as classes. O método de Random Forest corrige a questão de overfitting da base de treinamento na aplicação nos modelos de Árvores de Decisão.

As Árvores de Decisão são uma família popular de métodos para tarefas de classificação como análise de Churn, e seguramente, “pegando carona” nessa popularidade o método Random Forest acaba sendo bem utilizado também. No entanto, as Árvores de Decisão possuem certa instabilidade. Em particular, as árvores que são muito profundas tendem a aprender padrões altamente irregulares, que se refere, conforme mencionado, ao overfitting do conjunto de treinamento devido a baixa polarização e alta variância. Os modelos de Random Forest são uma forma de média de múltiplas Árvores de Decisão, formados de diferentes partes de um mesmo conjunto de treinamento, com o objetivo de reduzir a variância e melhorar o problema do overfitting. Por outro lado, para que se tenha este ganho no desempenho final, perde-se um pouco a interpretação do modelo, o que é especialmente negativo para a análise de Churn.

O algoritmo de treinamento do modelo aplica a técnica geral de agregação “bootstrap”, ou “bagging”, que é um método de modelos combinados, acrescido de uma outra característica explicada mais adiante.

Dado um conjunto de treinamento $X = x_1, \dots, x_n$ com respostas $Y = y_1, \dots, y_n$, faz-se o ensacamento (bagging) repetidamente (B vezes) selecionando uma amostra aleatória com substituição, do conjunto de treinamento e se ajusta árvores para estas amostras:

Para $b = 1, \dots, B$:

1. Amostra, com substituição, n exemplos de treinamento de X, Y ; são denomi-

nados X_b, Y_b .

2. Treinar uma Árvore de Decisão ou regressão, f_b , em X_b, Y_b .

Após o treinamento, as previsões para novas amostras x' , podem ser feitas através da média das previsões de todas as árvores de regressão individuais sobre x' :

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x')$$

ou tomando a moda no caso de Árvores de Decisão.

Este procedimento leva a um melhor desempenho do modelo porque diminui a variância do modelo, sem aumentar a tendência, conforme objetivo inicial mencionado. Isto significa que enquanto as previsões de uma única Árvore de Decisão são altamente sensíveis ao ruído no seu conjunto de treinamento, a média de muitas árvores não é, contanto que as árvores não sejam correlacionadas.

O número de amostras de árvores, B , é um parâmetro livre. Tipicamente, algumas centenas até vários milhares de árvores são usados, dependendo do tamanho e da natureza do conjunto de treinamento. Um número ideal de árvores B pode ser encontrado utilizando validação cruzada ou observando o erro de predição médio em cada amostra x_i de treinamento, usando apenas as árvores que não tinham x_i em sua amostra de bootstrap. Desta forma o erro tende a se estabilizar após um determinado número de árvores tenham sido ajustadas.

O procedimento acima descreve então como mencionado o algoritmo original para ensacamento (bagging) de árvores. O modelo de Random Forest propriamente dito difere em apenas um detalhe deste procedimento geral: ele usa um algoritmo de aprendizagem de árvore modificado, que seleciona a cada ramificação de candidato ao longo do processo de aprendizagem, um subconjunto aleatório de atributos. Este processo pode ser também conhecido como “ensacamento de atributos” (feature bagging). A razão para isso é a correlação das árvores em uma amostra de bootstrap comum: se um ou mais atributos são preditores muito fortes da variável de resposta (de saída), estas características serão selecionadas em muitas árvores B , fazendo com que elas se tornem correlacionadas entre si.

2.4.6 Outras Técnicas Encontradas

Aqui, como foi explicado no início da seção, foi criado uma subseção única, onde se dará continuidade as descrições de mais algumas técnicas analíticas de Churn encontradas no trabalho, que por critério de frequência de ocorrência, apareceram menos que as demais cinco já descritas, mas pelo menos duas vezes ao longo da pesquisa.

Vale ressaltar por hora, e posteriormente na seção de Metodologia voltaremos ao assunto, que algumas técnicas expostas nesta subseção, em geral, não são classificadas como técnicas de Machine Learning, porém as mesmas foram contempladas nessa pesquisa.

Métodos Combinados (Ensemble)

Em estatística e Machine Learning, Métodos Combinados (Ensemble) consiste em usar vários algoritmos de aprendizagem para obter um melhor desempenho preditivo do que poderia ser obtido a partir de qualquer um dos algoritmos de aprendizagem constituintes isoladamente. O termo Combinado (Ensemble) é geralmente reservado para os métodos que geram múltiplas hipóteses usando o mesmo método base de aprendizado que difere do termo mais amplo, de múltiplos sistemas classificadores, que também cobre a questão de multiplicidade, porém sem serem induzidos pelo mesmo método base de aprendizado.

Entre estes métodos tem-se:

Bootstrap aggregating (bagging), envolve ter cada modelo no conjunto de decisão final com peso igual de decisão, com objetivo de melhorar a variância do modelo. A técnica se utiliza de retiradas de subconjuntos aleatórios do conjunto de treinamento. Como exemplo, já mencionado neste trabalho, o algoritmo de Random Forest combina Árvores de Decisão.

Boosting, envolve gradualmente construir um conjunto através do treinamento de cada nova instância do modelo, baseando-se nas instâncias de treinamento classificadas erroneamente pelos modelos anteriores. Um exemplo e talvez a implementação mais comum de Boosting é o algoritmo AdaBoost.

Naive Bayes

Em aprendizagem de máquina, classificadores Naive Bayes, são uma família de classificadores probabilísticos simples com base na aplicação de teorema de Bayes com fortes (ingênuas) premissas de independência entre as características. Para alguns tipos de modelos de probabilidade, classificadores Naive Bayes podem ser treinados de forma muito eficiente em um ambiente de aprendizagem supervisionada.

De maneira geral, Naive Bayes é um modelo de probabilidade condicional em que dado uma instância de um problema a ser classificada, representado por um vetor $\mathbf{x} = (x_1, \dots, x_n)$ que representa algumas características n (variáveis independentes), ele atribui a esta instância probabilidades:

$$p(C_k|x_1, \dots, x_n)$$

para cada um dos possíveis resultados K ou classes.

O problema com a formulação anterior é que, se o número de funcionalidades n for grande ou se uma característica puder assumir um grande número de valores, então tal modelo baseando em tabelas de probabilidade é inviável. Portanto, utilizando o teorema de Bayes, para que o problema seja mais bem tratado, a probabilidade condicional pode ser decomposta como:

$$p(C_k|\mathbf{x}) = \frac{p(C_k) p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

O que seria apontado por alguns estudos como uma vantagem do modelo de Naive Bayes, é que ele requer apenas uma pequena quantidade de dados de treino para estimar os parâmetros necessários para a classificação.

Data Mining Evolutivo

Data Mining Evolutivo ou Data Mining Genético são termos genéricos para qualquer técnica de mineração de dados usando algoritmos evolucionários e pode ser utilizado em qualquer cenário de predição à base de classificação. Os algoritmos evolutivos para o trabalho de mineração de dados funcionam criando uma série de regras aleatórias a serem confrontadas com um conjunto de dados de treinamento. As regras que se encaixarem mais estreitamente aos dados são selecionadas. Este processo funciona de forma iterativa.

Análise de Sobrevivência

A análise de sobrevivência é um ramo da estatística para analisar a duração esperada de tempo até que um ou mais eventos aconteçam, como a morte nos organismos biológicos, falhas em sistemas mecânicos, e sobretudo no tema em questão, a saída ou cancelamento de um cliente de uma base de assinantes da empresa, o Churn. A análise de sobrevivência tenta responder a perguntas como: qual é a proporção de uma população que vai sobreviver após um determinado período de tempo? Dos que sobrevivem, a que taxa eles vão morrer ou não? Pode causas múltiplas de morte ou falha ser levado em conta? Como circunstâncias ou características particulares podem aumentar ou diminuir a probabilidade de sobrevivência?

Para responder a essas perguntas, é necessário definir “tempo de vida”. Em alguns casos isso pode estar bem definido, em outros nem tanto. De modo mais geral, análise de sobrevivência envolve a modelagem de tempo para os dados de eventos. Neste contexto, a morte ou o fracasso é considerado um “evento” na literatura de

análise de sobrevivência. A modelagem pode ser feita de diversas formas conforme a finalidade da mesma, como exemplos:

- Para descrever os tempos de sobrevivência de membros de um grupo
 - Tabelas de vida
 - As curvas de Kaplan-Meier
 - Função de sobrevivência
 - Função de risco
- Para comparar os tempos de sobrevivência de dois ou mais grupos
 - Teste log-rank
- Para descrever o efeito das variáveis categóricas e quantitativas sobre a sobrevivência
 - Regressão de Cox
 - Modelos de sobrevivência paramétricos
 - Árvores de sobrevivência
 - Random Forest de sobrevivência

Análise de Redes Sociais

Análise de redes sociais é o processo de investigação de estruturas sociais através do uso de teorias de rede e grafo. Ela caracteriza as estruturas em rede em termos de nós (atores individuais, pessoas ou coisas dentro da rede) e os laços ou arestas (relações ou interações) que os conectam. Exemplos de estruturas sociais comumente visualizadas através de análise de redes sociais incluem redes de mídia social, redes de amizades e conhecidos, grafos de colaboração, parentesco e transmissões de doenças. Existem métricas embutidas na análise destas redes referentes a conexões, distribuição e segmentação da rede que auxiliam na análise de uma forma quantitativa.

K-NN

Em reconhecimento de padrões, o K-ésimo Vizinho Mais Próximo (ou K-NN) é um método não paramétrico em geral utilizado para classificação e regressão. Em

ambos os casos, a entrada consiste nos k exemplos de treinamento mais próximos dentro do espaço de atributos. A saída depende se o algoritmo é utilizado para classificação ou regressão.

No caso de classificação, como no estudo de Churn, a saída é uma associação de classe. Um objeto é classificado pelo voto da maioria de seus vizinhos, isto é, objeto é associado à classe mais comum entre os seus k vizinhos mais próximos (k é um inteiro positivo, tipicamente pequeno). Se $k = 1$, então o objeto é simplesmente atribuído à classe desse único vizinho mais próximo. Por outro lado, se $k = n$, sendo n o número total de vizinhos, este objeto será atribuído a classe dominante do conjunto de n objetos.

O algoritmo K-NN está entre os mais simples de todos os algoritmos de Machine Learning.

Modelos de Markov

Em teoria da probabilidade, um modelo de Markov é um modelo estocástico utilizado para modelar aleatoriamente sistemas onde se presume que os estados futuros dependem apenas do estado atual e não dos eventos que ocorreram antes. (propriedade de Markov). Geralmente, esta hipótese permite raciocínio e cálculo com o modelo que de outra forma seriam difíceis de resolver.

Redes Bayesianas

Rede Bayesiana, ou Rede de Bayes, ou modelo gráfico acíclico dirigido probabilístico é um modelo gráfico probabilístico (um tipo de modelo estatístico) que representa um conjunto de variáveis aleatórias e suas dependências condicionais através de um grafo acíclico dirigido (DAG). Por exemplo, uma rede Bayesiana poderia representar as relações de probabilidade entre doenças e sintomas. Dados os sintomas, a rede pode ser usada para calcular as probabilidades da presença de várias doenças.

3 Metodologia de Pesquisa

Os artigos foram inicialmente procurados através da ferramenta de busca do Google. Isto é, foram utilizadas as palavras “Data Mining”, “Machine Learning” e “Churn” de preferência em conjunto, nas combinações “Data Mining” + “Churn” e “Machine Learning” + “Churn” com objetivo de encontrar os artigos, pesquisas e documentos que tratassem especificamente do tema deste trabalho.

A ideia de utilização do Google seria ter um ponto de partida de busca de documentos, evitando um viés de busca em algum repositório ou periódico on line específicos, e também contornando eventuais necessidades de assinatura e autorizações de acesso aos documentos. Ou seja, não foi feito ou pago nenhuma permissão de acesso aos documentos utilizados na pesquisa, todos os artigos utilizados neste trabalho, foram encontrados para baixar de forma livre na web.

Logo, o primeiro passo foi colocar na ferramenta de busca do Google os termos “Data Mining” + “Churn” e “Machine Learning” + “Churn”. Com o retorno destas pesquisas foram sendo analisados os resumos de cada artigo e baixados aqueles que estavam em linha com os critérios de pesquisa deste trabalho. Um número inicial de mais ou menos 30 artigos foram baixados para serem posteriormente melhor analisados. O critério básico para a consideração do arquivo era tratar o problema de Churn, sem restrições de área de aplicação e sem restrições de ano de publicação, por alguma técnica já conhecida e aceita como de Data Mining ou Machine Learning conforme variação de nomenclaturas e definições já explicadas nas sessões anteriores deste trabalho. Não era necessariamente obrigatório que o artigo tratasse exclusivamente do problema de Churn, por exemplo, foram considerados artigos que tratavam de processos de CRM em geral, mas desde que houvesse alguma seção do trabalho tratando isoladamente a questão do Churn. Da mesma forma, também não foram descartados artigos que tivessem, simultaneamente, outros tipos de técnicas de tratamento do problema de Churn que não essencialmente de Machine Learning que estávamos em busca. E mais ainda, devido a dificuldade ou falta de uma classificação universal das técnicas, houve uma flexibilização neste critério e artigos com outras análises técnicas julgadas relevantes (sem um critério técnico de relevância definido neste caso), desde que o processo de análise técnica estivesse sendo explicado de forma clara e que com isso pudessem acrescentar valor ao trabalho, também não foram descartados. Há que se destacar ainda dentro desta etapa, que eventualmente, quando não se conseguia baixar o arquivo referente ao artigo encontrado, outras pesquisas on line eram tentadas, na esperança de encontrar o arquivo disponível para baixar em outro repositório alternativo àquele inicialmente encontrado. Os acessos foram feitos ou via conexão privada particular ou via conexão privada da FGV,

onde em alguns casos apresentaram liberação para repositórios antes bloqueados no acesso particular. Conforme critério estabelecido, os documentos não conseguidos cópia na íntegra para nosso repositório local foram descartados. Isso porque, todos os documentos conforme baixados e armazenados, posteriormente, passariam por uma segunda análise mais completa, além do que, seria um facilitador ter posse dos documentos para a extração dos dados de interesse da pesquisa. Nessa etapa mais alguns artigos também seriam descartados.

Uma primeira base de documentos de 20 artigos foi encontrada e armazenada localmente após essa primeira rodada de pesquisa. Estes artigos então passaram por uma classificação e parametrização, onde para cada artigo foram extraídos os seguintes atributos da tabela abaixo e tabelados em uma planilha única.

Tabela 1: Tabela de Atributos dos Artigos

Atributo	Descrição
ID	Neste campo foi dado um número de registro para cada artigo.
Nome do Artigo	Nome do Artigo.
Ano	Ano de publicação do artigo.
Repositório	Nome do repositório onde foi encontrado o artigo para download.
Periódico	Nome do periódico onde foi publicado o artigo.
Autor	Nome dos autores do artigo.
ALIAS autor	Nome, apelido criado para identificação mais pratica do artigo.
Resumo	Resumo (Abstract) na íntegra do artigo.
Area	Área de aplicação da análise de Churn do artigo.
Técnica	Técnica Análítica (ML, DM, Outras) aplicada ao Churn no artigo.
Referências	Todas as referências utilizadas pelo artigo.
ALIAS autor re-ferência	Apelido criado para identificação mais pratica de cada artigo nas referências.
OBS	Eventual observação geral utilizada.

Nota-se na tabela acima um atributo chamado “ALIAS autor referência”. Este atributo foi criado, para que através do mesmo, fosse possível, utilizando uma ferramenta de software (Excel), contar o número de citações que determinado artigo tinha

recebido. E assim sendo, foi feito um ranking entre este grupo de 20 artigos baixados, pelo número de citações recebidas por cada artigo. Desta forma, estabeleceu-se que para os artigos com o número de citações maior ou igual a 10% do número total de artigos naquela determinada rodada de pesquisa, caso o artigo já não estivesse na base local, o mesmo seria procurado na internet reiniciando assim o fluxo de busca e análise dos artigos conforme já descrito acima. Assim, por exemplo, para esta primeira base de artigos que continha 20 artigos, cada um com duas ou mais citações, seria candidato a busca na internet. Perceba que desta forma a base de artigos possuía um processo cíclico de alimentação de artigos, com a qual a base foi crescendo. Ressaltamos aqui, que aportes externos a este processo também foram aceitos. O que isso quer dizer? Que ao longo da busca por determinados artigos recomendados pelos critérios estabelecidos para o ranking, se outros artigos pertinentes eram casualmente encontrados, estes eram aceitos a seguir o fluxo do processo de aquisição. E desta forma seguiu-se por mais algumas rodadas de atualizações, até que nossos critérios de parada fossem atingidos.

Os critérios de parada arbitrados foram os seguintes:

1. Número total de artigos entre 50 -100. Quantidade estabelecida pela equipe, em função do prazo do trabalho, força de trabalho empregada e por entender que já seria uma quantidade significativa para os propósitos, desde que também atingidos os demais critérios a seguir.
2. Todos os artigos, relevantes, e que pudessem ser baixados, com o índice de citações de maior ou igual a 10% do total dos artigos, já coletados.
3. Saturação. Critério estabelecido de forma qualitativa, isto é, sem valor numérico definido, que se refere ao alto grau de dificuldade de encontrar novos artigos aderentes ao tema e critérios estabelecidos a partir de uma dada quantidade de artigos já coletados.

Tendo em vista que ao longo do processo, a primeira análise era feita baseando-se no resumo do artigo e de forma on line, não foi coletado o dado de quantos artigos foram percorridos. Estima-se que algo próximo ao dobro do número de artigos que ao final formaram a base da pesquisa, tenham sido verificados neste processo. De onde um total de 102 documentos foram baixados e destes um total de 80 artigos foram aceitos e reunidos para o trabalho de classificação final.

Abaixo é apresentado um diagrama do processo de pesquisa descrito ao longo do texto acima. E logo em seguida a tabela com os 80 artigos coletados para o trabalho, através desta metodologia apresentada. Observa-se que apenas os atributos: “ID” ,

“Nome do artigo”, “Citações” e “Autor”, sendo este último apresentado como formato de referência no texto, foram contemplados na tabela por motivos de adequação a visualização da tabela.

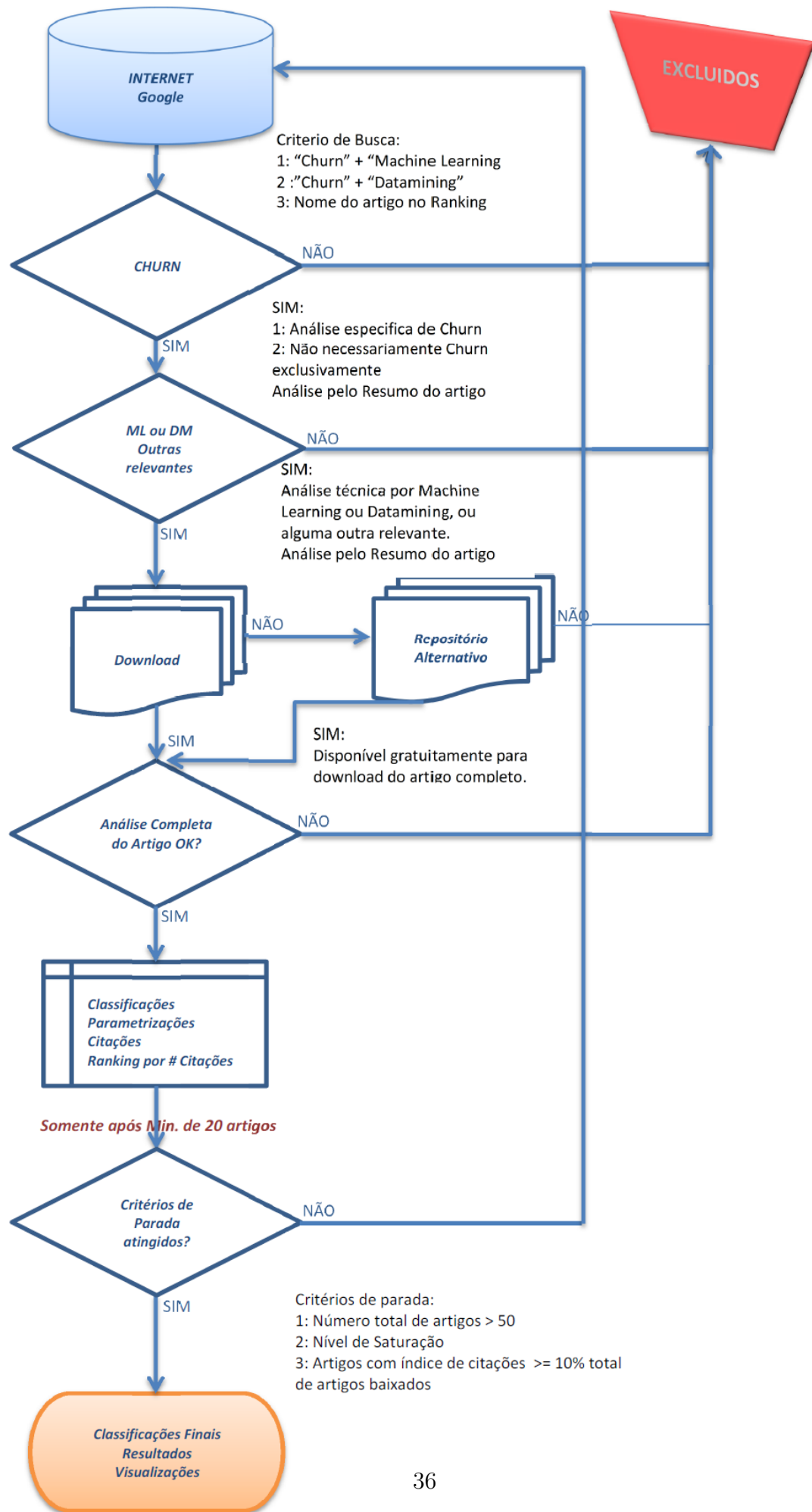


Tabela 2: Tabela de Artigos

ID	Nome do Artigo	Citações	Autor(Referência)
1	Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network	0	Castanedo et al. (2014)
2	Customer Churn prediction using improved balanced random forests	7	Xie et al. (2009)
3	Model of Customer Churn Prediction on Support Vector Machine	0	Xia and Jin (2008)
4	Churn detection in large user networks	0	Deri and Moura (2014)
5	A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services	0	Sharma, Panigrahi and Kumar (2013)
6	Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques	22	Coussement and Van den Poel (2008)
7	A Hybrid Churn Prediction Model in Mobile Telecommunication Industry	0	Olle and Cai (2014)
8	Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry	21	Mozer et al. (2000)
9	Applying data mining to telecom Churn management	25	Hung, Yen and Wang (2006)
10	Churn Prediction using Complaints Data	5	Hadden et al. (2006)
11	Customer Churn prediction - a case study in retail banking	0	Mutanen, Ahola and Nousiainen (2006)
12	Prediction of Churn behavior of bank customers using Data Mining tools	1	Prasad and Madhavi (2012)
13	Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining	1	Kirui et al. (2013)
14	Predicting Churn in mobile telecommunications industry	0	TODERAN (N.d.)

15	Predicting Customer Retention and Profitability by Using Random Forests and Regression Forests Techniques	0	Larivière and Van den Poel (2005)
16	Application of data mining techniques for customer lifetime value parameters: a review	0	Aeron, Kumar and Janakiraman (2010)
17	Customer Churn Prediction Using Improved One-Class Support Vector Machine	7	Zhao et al. (2005)
18	Employee Churn prediction	0	Saradhi and Palshikar (2011)
19	Winning the KDD Cup Orange Challenge with Ensemble Selection	0	Niculescu-Mizil et al. (2009)
20	Predicting customer behaviour: The University of Melbourne's KDD Cup report	0	Miller et al. (2009)
21	A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction	19	Au, Chan and Yao (2003)
22	An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry	13	Hwang, Jung and Suh (2004)
23	An SVM based Churn Detector in Prepaid Mobile Telephony	6	Archaux, Martin and Khenchaf (2004)
24	Data mining techniques for customer relationship management	3	Rygielski, Wang and Yen (2002)
25	Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models	15	Neslin et al. (2006)
26	Goal-oriented sequential pattern for network banking Churn analysis	10	Chiang et al. (2003)
27	A new feature set with new window techniques for customer Churn prediction in land-line telecommunications	3	Huang et al. (2010)
28	A recommender system to avoid customer Churn: A case study	0	Wang et al. (2009)

29	Building comprehensible customer Churn prediction models with advanced rule induction techniques	3	Verbeke et al. (2011)
30	Churn models for prepaid customers in the cellular telecommunication industry using large data marts	2	Owczarczuk (2010)
31	Computer Assisted Customer Churn Management: State-Of-The-Art and Future Trends	9	Hadden et al. (2007)
32	Customer attrition analysis for financial services using proportional hazard models	21	Van den Poel and Larivière (2004)
33	Customer Churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry	3	Ahn, Han and Lee (2006)
34	Customer Churn prediction by hybrid neural networks	4	Tsai and Lu (2009)
35	Customer Churn prediction in telecommunications	5	Huang, Kechadi and Buckley (2012)
36	Determinants of customer loyalty in the wireless telecommunications industry	1	Eshghi, Haughton and Topi (2007)
37	Determinants of subscriber Churn and customer loyalty in the Korean mobile telephony market	14	Kim and Yoon (2004)
38	Handling class imbalance in customer Churn prediction	10	Burez and Van den Poel (2009)
39	Improved marketing decision making in a customer Churn prediction context using generalized additive models	5	Coussement, Benoit and Van den Poel (2010)
40	Investigating the role of product features in preventing customer Churn, by using survival analysis and choice modeling:The case of financial services	9	Larivière and Van den Poel (2004)
41	Modeling Churn using customer lifetime value	0	Glady, Baesens and Croux (2009)
42	Multi-objective feature selection by using NSGA-II for customer Churn prediction in telecommunications	2	Huang, Buckley and Kechadi (2010)

43	New insights into Churn prediction in the telecommunication sector: A profit driven data mining approach	4	Verbeke et al. (2012)
44	Predicting customer Churn in mobile networks through analysis of social groups	1	Richter, Yom-Tov and Slonim (2010)
45	Variable selection by association rules for customer Churn prediction of multimedia on demand	0	Tsai and Chen (2010)
46	Application of data mining techniques in customer relationship management: A literature review and classification	4	Ngai, Xiu and Chau (2009)
47	Customer relationship management research(1992-2002) An academic literature review and classification	1	Ngai (2005)
48	Targeting customers with statistical and data-mining techniques	5	Drew et al. (2001)
49	Customer Churn Prediction in Telecommunication A Decade Review and Classification	0	Hashmi, Butt and Iqbal (2013)
50	Domain Knowledge Integration in Data Mining using Decision Tables: Case Studies in Churn Prediction	4	Lima, Mues and Bae-sens (2009)
51	An integrated data mining and behavioral scoring model for analyzing bank customers	4	Hsieh (2004)
52	Defection Detection: Improving Predictive Accuracy of Customer Churn Models	5	Neslin et al. (2004)
53	Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market	7	Gerpott, Rams and Schindler (2001)
54	The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services	5	Kim, Park and Jeong (2004)
55	Bagging and Boosting Classification Trees to Predict Churn	15	Lemmens and Croux (2006)
56	Modeling Customer Lifetime Value	4	Gupta et al. (2006)

57	A comparison of Machine Learning techniques for customer Churn prediction	0	Vafeiadis et al. (2015)
58	A novel decision rules approach for customer relationship management of the airline market	1	Liou (2009)
59	An effective hybrid learning system for telecommunication Churn prediction	1	Huang and Kechadi (2013)
60	An empirical evaluation of rotation-based ensemble classifiers for customer Churn prediction	6	De Bock and Van den Poel (2011)
61	An extended support vector machine forecasting framework for customer Churn in e-commerce	1	Yu et al. (2011)
62	Applying Bayesian Belief Network approach to customer Churn analysis: A case study on the telecom industry of Turkey	3	Kisioglu and Topcu (2011)
63	Churn management optimization with controllable marketing variables and associated management costs	0	Kim, Lee and Johnson (2013)
64	Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies	0	Idris, Rizwan and Khan (2012)
65	Combined rough set theory and flow network graph to predict customer Churn in credit card accounts	0	Lin, Tzeng and Chin (2011)
66	CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services	10	Burez and Van den Poel (2007)

67	Customer event history for Churn prediction: How long is long enough?	1	Ballings and Van den Poel (2012)
68	Dynamic Churn prediction framework with more effective use of rare event data: The case of private banking	0	Ali and Arıttürk (2014)
69	Estimating the effect of word of mouth on Churn and cross-buying in the mobile phone market with Markov logic networks	0	Dierkes, Bichler and Krishnan (2011)
70	Genetic algorithm based neural network approaches for predicting Churn in cellular wireless network services	5	Pendharkar (2009)
71	GSM Churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system	1	Karahoca and Karahoca (2011)
72	Improved Churn prediction in telecommunication industry by analyzing a large network	2	Kim, Jun and Lee (2014)
73	Improved Churn prediction in telecommunication industry using Data Mining techniques	0	Keramati et al. (2014)
74	Improving customer attrition prediction by integrating emotions from client-company interaction emails and evaluating multiple classifiers	2	Coussement and Van den Poel (2009)
75	Improving customer retention in financial services using kinship network information	0	Benoit and Van den Poel (2012)
76	Managing B2B customer Churn, retention and profitability	0	Jahromi, Stakhovych and Ewing (2014)
77	Profit-based feature selection using support vector machines General framework and an application for customer retention	0	Maldonado et al. (2015)
78	Reconciling performance and interpretability in customer Churn prediction using ensemble learning based on generalized additive models	0	De Bock and Van den Poel (2012)
79	Sequential manifold learning for efficient Churn prediction	0	Kim and Lee (2012)
80	Turning telecommunications call details to Churn prediction- a data mining approach	27	Wei and Chiu (2002)

4 Apresentação dos Resultados

Após a coleta dos artigos conforme detalhadamente explicado nas seções anteriores e armazenamento na base de dados, isto é, em uma única e grande planilha, foi feita uma exploração dos dados, e outras planilhas e campos de atributos auxiliares foram criados conforme necessidade, a fim de extrairmos informações dos dados coletados. Nesta seção serão então apresentadas as visualizações dos resultados obtidos nesta pesquisa para os 80 artigos da base de dados. Nas subseções seguintes, serão expostas as classificações e interpretações de acordo com 6 diferentes dimensões a saber: Técnicas de Machine Learning empregadas, Repositórios, Periódico, Ano de Publicação, Área de Aplicação e Citações.

4.1 Classificação por Técnicas de Machine Learning Empregadas

A figura 10 mostra a classificação dos artigos pelas técnicas de Machine Learning empregadas. Para um total de 161 ocorrências de técnicas encontradas, Árvores de Decisão lidera com 19% (31) de utilização, seguida de Regressão Logística com 16% (25) e Redes Neurais com 14% (22). Perceba que temos um total de 161 ocorrências de técnicas encontradas em um universo de 80 artigos selecionados, o que mostra que os artigos têm em média mais de uma técnica sendo utilizada. Tem-se ainda no gráfico uma classe “outras” com também 14% (22) que representam técnicas que apareceram, cada uma, uma única vez neste conjunto de artigos.

Os resultados apresentaram claramente 3 técnicas em destaque nesta pesquisa: Árvores de Decisão, Regressão Logística e Redes Neurais. Acreditamos que este destaque deve-se ao fato da tradição e maturidade dos estudos destas técnicas. Acreditamos num maior destaque das duas primeiras técnicas pela simplicidade de aplicação das mesmas e por serem eficientes para análises qualitativas e posteriormente de fácil interpretação dos resultados, facilitando a compreensão de profissionais menos especialistas nos algoritmos e mais voltados para o negócio. Porém estes tipos de técnicas não garantem, em geral, um desempenho e precisão tão bons quanto a técnica de Redes Neurais, principalmente se o conjunto de dados for muito grande, com muitas dimensões, ou não linear. Com isso, embora com a desvantagem de ser uma técnica um pouco mais complexa, sem o poder de explicação das demais, Redes Neurais também apresentou grande destaque provavelmente em função da eficiência de seus resultados. Vale apenas ainda ressaltar a aparição da técnica de Random Forest entre o grupo das 5 primeiras técnicas com maior ocorrência. Isso porque, ainda na seção de teoria, foi mencionado que a agregação de várias Árvores de Decisão implicaria em

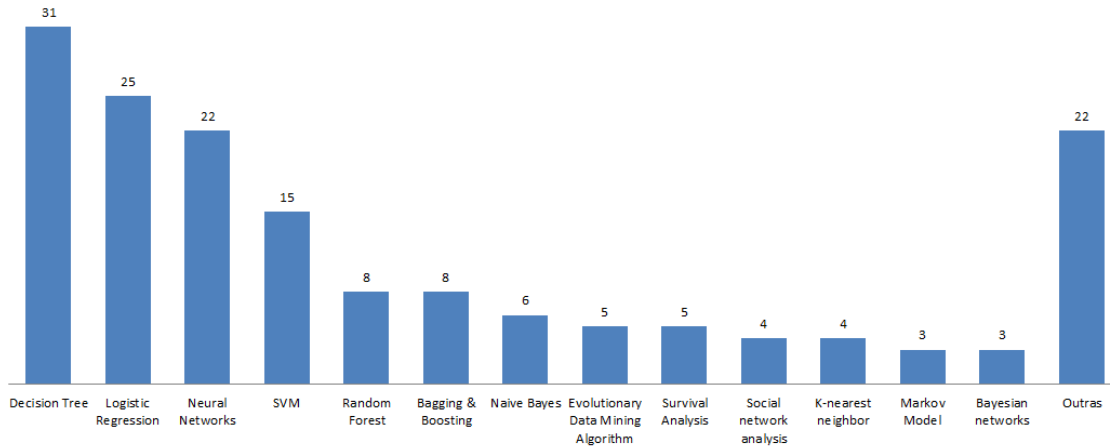


Figura 10: Técnicas de Machine Learning.

uma perda de interpretação dos resultados ao final do modelo. Tendo dito isto, pelo resultado encontrado, podemos assumir então que o benefício trazido ao desempenho do modelo por esta agregação parece superar o fato da perda de interpretação. De fato, é difícil saber apriori qual a técnica que apresentará melhor desempenho com dados disponíveis, e por essa razão, podemos perceber pelos números apresentados, que uma solução encontrada nos diversos estudos pesquisados, é a aplicação de diferentes técnicas simultaneamente para o problema. Trata-se de uma solução eficaz seja pela abrangência de diferentes características das diferentes técnicas empregadas ou pela exaustão na busca de resultados cada vez melhores.

Para informação e por não pretender excluir novas interpretações, é listado aqui, conforme nomenclatura dos respectivos artigos, as 22 outras técnicas que compuseram a classe “outras” em nosso julgamento. São elas: *Deep Learning*; *Fuzzy Logic*; *Ensemble Selection*; *Nonparametric Feature Selection*; *Sequential Analysis*; *AntMiner+ ALBA*; *Linear regression*, *Fisher linear discriminant analysis*; *Linear Classifications*; *Tetrad methodology and Structural Equation Model (SEM)*; *Association Rules*; *Basic Stat*; *Variable Consistency Dominance-based Rough Set Approach (VC-DRSA)*; *K-means*, *rule inductive technique (FOIL)*; *Rotation Forest and RotBoost*; *Partial least square (PLS) optimization method*; *Rough set theory*, *Flow network graph*; *Cox regression*; *Adaptive Neuro Fuzzy Inference System (ANFIS)*; *Ensemble classification*; *Generalized additive models (GAMs)*; *Manifold learning*; *LISREL*.

4.2 Classificação por Área de aplicação

A figura 11 mostra a classificação dos artigos por área de Aplicação. A grande área de concentração encontrada para este tipo de estudo foi a área de Telecomunicações representando 60% do total de 82 ocorrências de áreas encontradas. Seguida pela área financeira com 16% , divididos entre 10% de Bancos e 6% de outras áreas que não Bancos, como operadoras de cartões de crédito por exemplo ou que no artigo não foi explicitamente mencionada, porém informado como sendo do setor financeiro.

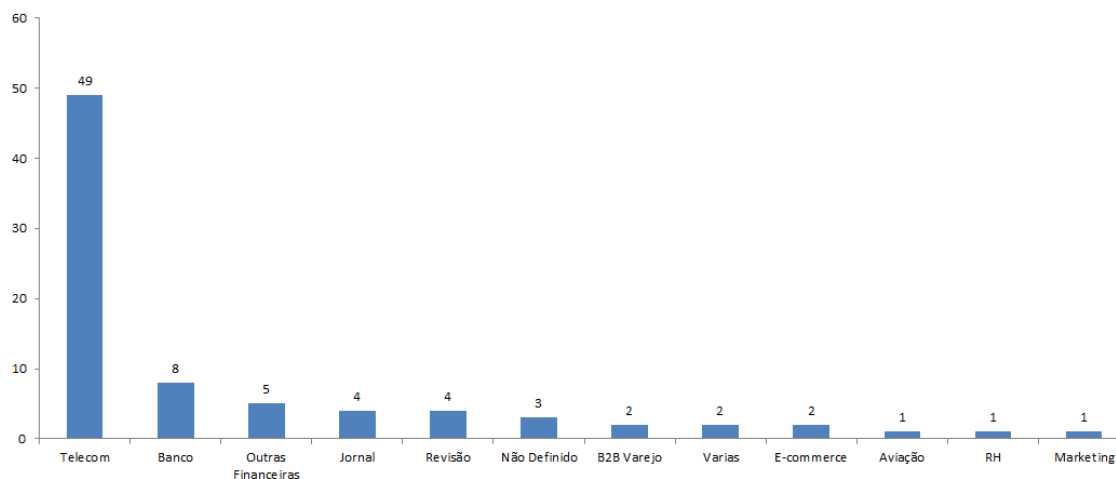


Figura 11: Classificação por Área de aplicação.

Nota-se pelo resultado a grande concentração deste tipo de estudo na área de Telecomunicações. cremos que este resultado era esperado devido a combinação de fatores anteriormente citados no trabalho, como alta competitividade no setor, grande volume de dados gerados através das massas de ligações telefônicas e conexões de dados dos milhões de usuários espalhados pelas redes. Além da aquisição dos dados, que embora atualmente esteja cada vez mais burocratizada e sigilosa, ainda sim, parece ser menos complexa e mais acessível que da área financeira, a segunda com maior concentração de estudos detectada.

Observa-se também, assim como na classificação por técnica, que alguns artigos apresentaram estudos para mais de uma área de aplicação ao mesmo tempo e por isso se obteve um total de ocorrências superior ao total de 80 documentos avaliados. Dois artigos ainda apresentaram avaliações de técnicas em diversas áreas e, portanto, não se caracterizando que o artigo era um estudo específico de uma determinada área

e assim optou-se denominar esta área de aplicação como “Várias”.

4.3 Classificação por Ano de Publicação

A figura 12 mostra a classificação dos artigos por ano de Publicação. Observa-se que os primeiros artigos encontrados sobre o tema datam do ano 2000. Observamos ainda que em nossa pesquisa os anos de 2004 e 2009 foram os que mais apresentaram materiais sobre o assunto com 11,3% e 15% respectivamente do total de publicações. E o ano de 2000, com apenas 1 artigo, foi o que menos apresentou publicações do assunto, representando 1,3% do total. Ainda no gráfico, foram representadas duas linhas auxiliares de visualização que são: em vermelho uma linha de tendência linear, e em azul claro uma linha de percentual acumulado.

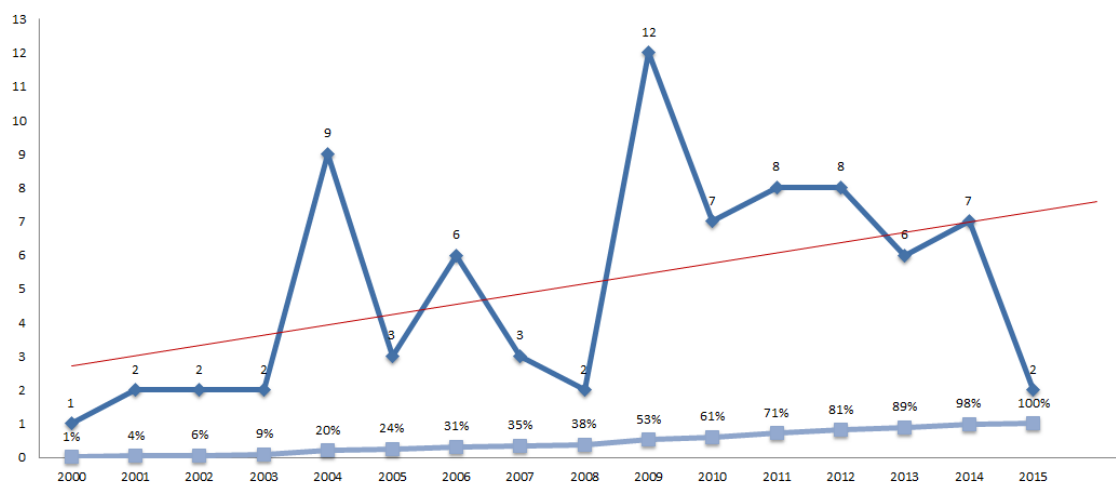


Figura 12: Classificação por Ano de Publicação.

Recordando que não houve restrições de datas como critério de decisão em nossa pesquisa, tende-se a concluir que o ano 2000 pode estar bem próximo do ano em que realmente surgiram os primeiros estudos do assunto. De forma mais ampla, esta hipótese pode se referir aos primeiros estudos de análise de Churn (menos provável), ou estudos utilizando especificamente os termos Data Mining e/ou Machine Learning para análise de Churn, ou ainda, as primeiras publicações oficiais sobre o tema. Embora o ano de 2004 tenha apresentado o segundo maior número de publicações encontradas, percebe-se que a partir do ano de 2009 é que houve uma concentração

maior neste tema de pesquisa. Isso pode ser percebido na figura pelas linhas auxiliares tanto pela linha de tendência linear que demonstra, embora com variações, a tendência de crescimento compreendida entre os anos de 2000 a 2015, quanto no gráfico de percentual acumulado, mostrando um descolamento maior do eixo, isto é, um crescimento mais acelerado a partir do ano de 2009.

Acreditamos que, entre outros fatores, alguns possam explicar essa tendência, tais como:

- Aumento na competitividade dos setores levando a uma necessidade de estudos analíticos mais detalhados agregando valor às estratégias competitivas.
- Evolução da tecnologia da internet e suas conexões, dos computadores e dispositivos móveis pessoais, aumentando assim a geração e disponibilidade de dados para tais estudos analíticos.
- Evolução também na tecnologia para armazenamento e processamento de maiores quantidades de dados.
- Amadurecimento e disseminação do conhecimento e técnicas utilizadas nas pesquisas.

Observam-se ainda valores perto do mínimo, fora da tendência discutida acima para os anos de 2008 e 2015. Para o primeiro, realmente foi um resultado fora do esperado, de difícil ou sem explicação, ou talvez por falta de dados. O segundo caso, a não ser que o ano de 2015 realmente seja o início de uma tendência de baixa ou de saturação de estudos de Churn, e que só será possível essa constatação com futuras pesquisas como esta, entendemos que pode ser explicado pelo fato da pesquisa, no que se refere especialmente a coleta dos artigos, ter sido feita em 2015 e assim talvez não ter capturado os artigos mais recentes.

4.4 Classificação por Periódico

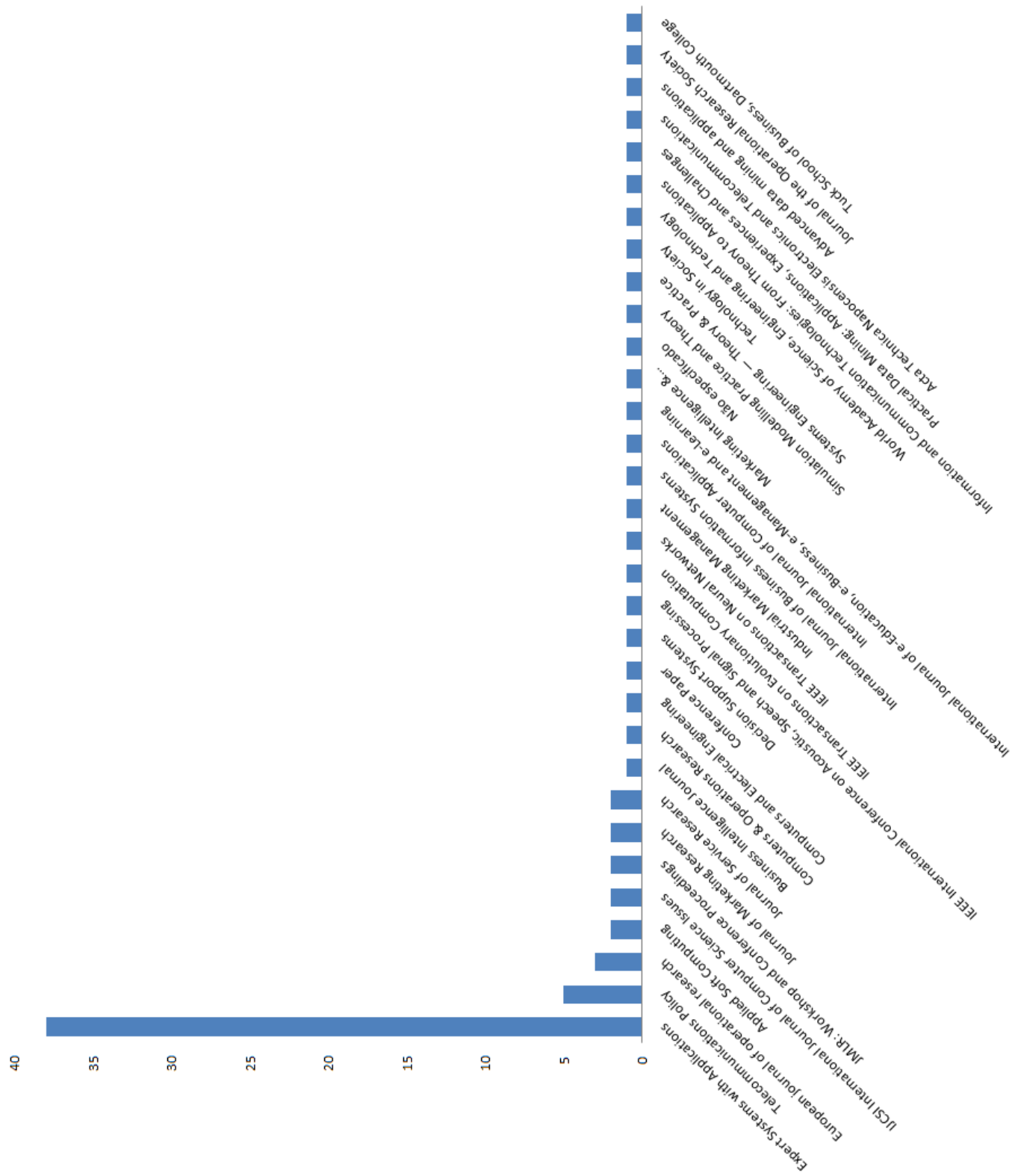
A figura 13 mostra a classificação dos artigos por periódicos científicos nos quais os artigos foram publicados. Um total de 32 diferentes periódicos apareceram na pesquisa, sendo destes, *Expert Systems with Applications* o que de longe mais apareceu com quase metade dos artigos utilizados no trabalho (48%). Ainda, *Telecommunications Policy* e *European Journal of Operational Research* tiveram percentuais mais expressivos com 6% e 4% respectivamente.

Tendo em vista a grande variedade de periódicos existentes, esperava-se uma distribuição mais equilibrada para esta classificação, o que não ocorreu devido à

predominância do periódico *Expert Systems with Applications*. Periódicos associados diretamente à área de telecomunicações também eram esperados que aparecessem com maior volume justamente por ser a área de aplicação com maior quantidade de estudos de Churn.

Uma análise da linha editorial do periódico para entender a reputação e características deste periódico, nos revelou que embora seja um periódico com alto volume de publicações de Machine Learning, o mesmo não consta entre os mais volumosos. Porém, o mesmo se define como: “um periódico internacional arbitrado cujo foco é a troca de informações relativas aos sistemas especialistas e inteligentes aplicados na indústria, governo e universidades em todo o mundo.”. O que talvez explique a aderência ao tema, e a concentração do mesmo, uma vez que não estamos estudando especificamente métodos e teorias de Machine Learning, mas sim uma aplicação específica importante para as áreas mencionadas na própria definição do periódico.

Embora se tenha tentando evitar esta concentração de um ou de poucos periódicos, percebemos que a metodologia utilizada no que se refere a buscar na internet os artigos mais citados em cada rodada de busca, pode ter levado a este viés de busca e consequentemente a esta concentração. Isso porque, em geral, possivelmente haja uma tendência de que artigos publicados em um periódico cite outros artigos publicados no mesmo periódico.



4.5 Classificação por Repositório

A figura 14 mostra a classificação dos artigos por Repositórios de onde os artigos foram encontrados. Percebe-se que 4 repositórios concentraram quase 80% dos artigos, são eles: ScienceDirect (44%), Elsevier (19%), Researchgate (10%) e IEEE(5%). Sendo ainda que o repositório ScienceDirect na verdade hoje é uma solução digital do portfólio Elsevier com sua coleção de textos eletrônicos de periódicos científicos de publicação Elsevier.

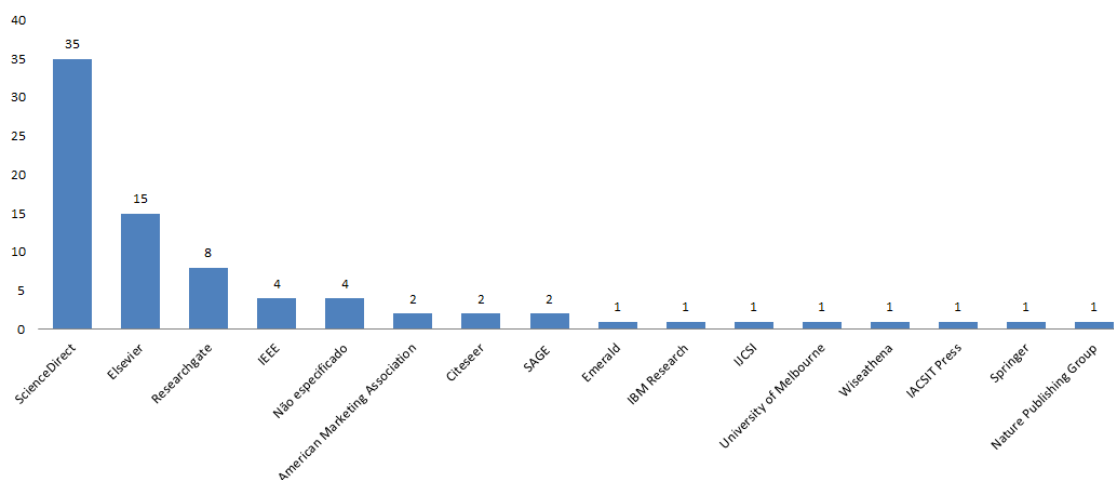


Figura 14: Classificação por Repositório.

Também se esperava, contudo, em função da metodologia utilizada onde a ideia era não privilegiar algum repositório específico, que estas fontes fossem mais bem distribuídas, o que não ocorreu. Porém, tendo em vista a concentração de periódicos encontrada na classificação anterior, isso poderia vir a explicar diretamente a concentração dos repositórios, já que determinados periódicos pertencem a determinados repositórios. Também não se pode descartar a inversão da causa e efeito justificada nestas duas últimas seções de classificação, isto é, que a concentração dos repositórios tenha causado a concentração dos periódicos. Encontramos ainda um total em torno de 5% de repositórios não identificados. Isso aconteceu provavelmente em função da busca não direcionada, onde algumas vezes baixou-se diretamente da pagina principal de busca do Google sem que esta informação fosse notada e tampouco constava no documento baixado.

4.6 Classificação por Ano & Técnicas de Machine Learning.

Nesta seção, combinamos duas dimensões de classificações, ou seja, a figura 15 mostra a classificação combinada de Ano de Publicação e Técnicas de Machine Learning. Foi utilizada uma técnica de visualização chamada “Tree Map”, onde neste caso, os anos estão representados pelas cores, as técnicas através dos rótulos com os próprios nomes e, a característica principal desta visualização, é que as áreas dos retângulos são proporcionais as suas quantidades, que neste caso são as quantidades de artigos.



Figura 15: Tree Map: Classificação por Ano & Técnicas de Machine Learning.

Por esta nova imagem, percebe-se e confirma-se o que foi mencionado em classificações anteriores, que existe uma predominância de material nos anos mais recentes. Isso pode ser visto pela correspondente predominância das cores com tons de azul sobre as cores de tons mais avermelhados ou brancos na figura. Por sua vez, internamente a cada região de cores pode-se notar a predominância das 3 técnicas de Machine Learning que mais apareceram no estudo, isto é, Árvores de Decisão, Regressão Logística e Redes Neurais.

Já na figura 16 repetiu-se o mesmo estudo combinado acima, porém com uma visualização mais clássica. Ou seja, na figura 16 é apresentada a classificação combinando os atributos Ano e Técnicas de Machine Learning, em um gráfico de Barras.

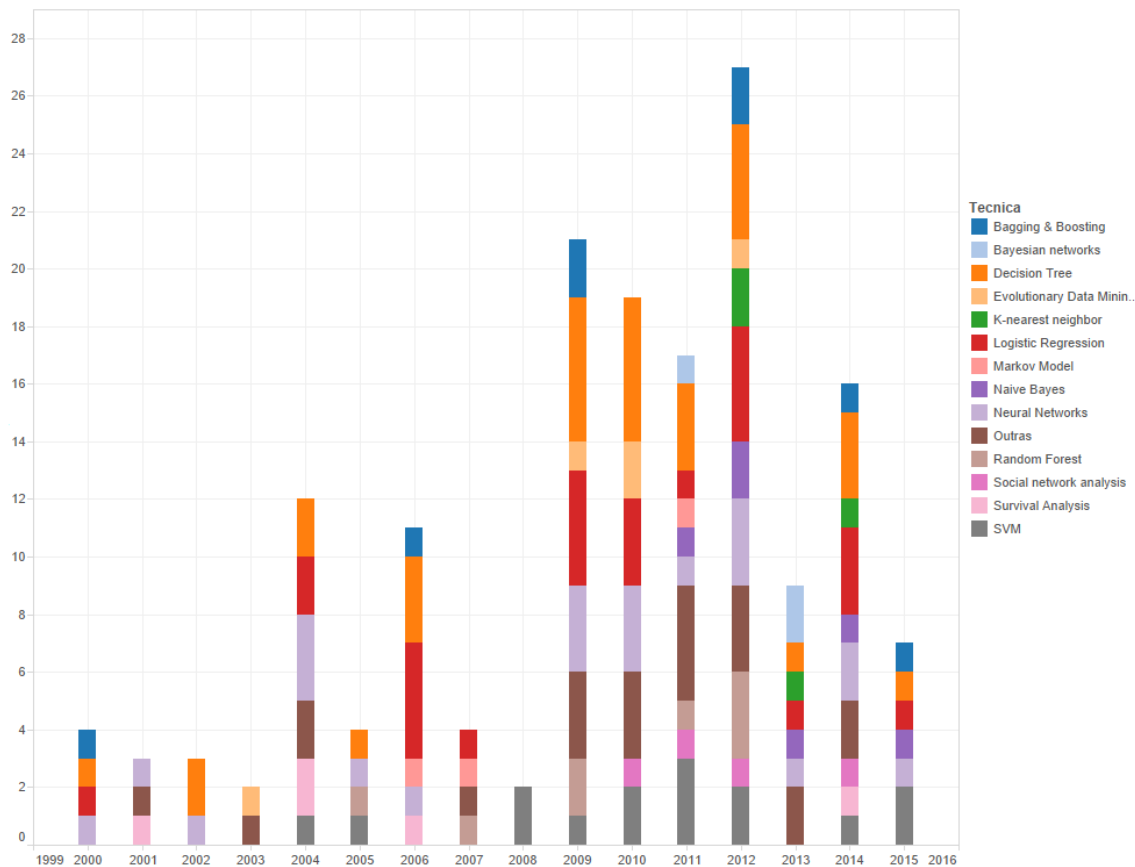


Figura 16: Barras: Classificação por Ano & Técnicas de Machine Learning.

E com isso, na figura 17, com esta mesma visualização em Barras, foi possível implementar um filtro e poder mostrar de maneira destacada apenas as 3 técnicas predominantes no estudo. Percebe-se com isso que exceto em dois anos do período encontrado, os anos de 2004 e 2008, pelo menos uma destas 3 técnicas aparecem em estudos nos demais anos.

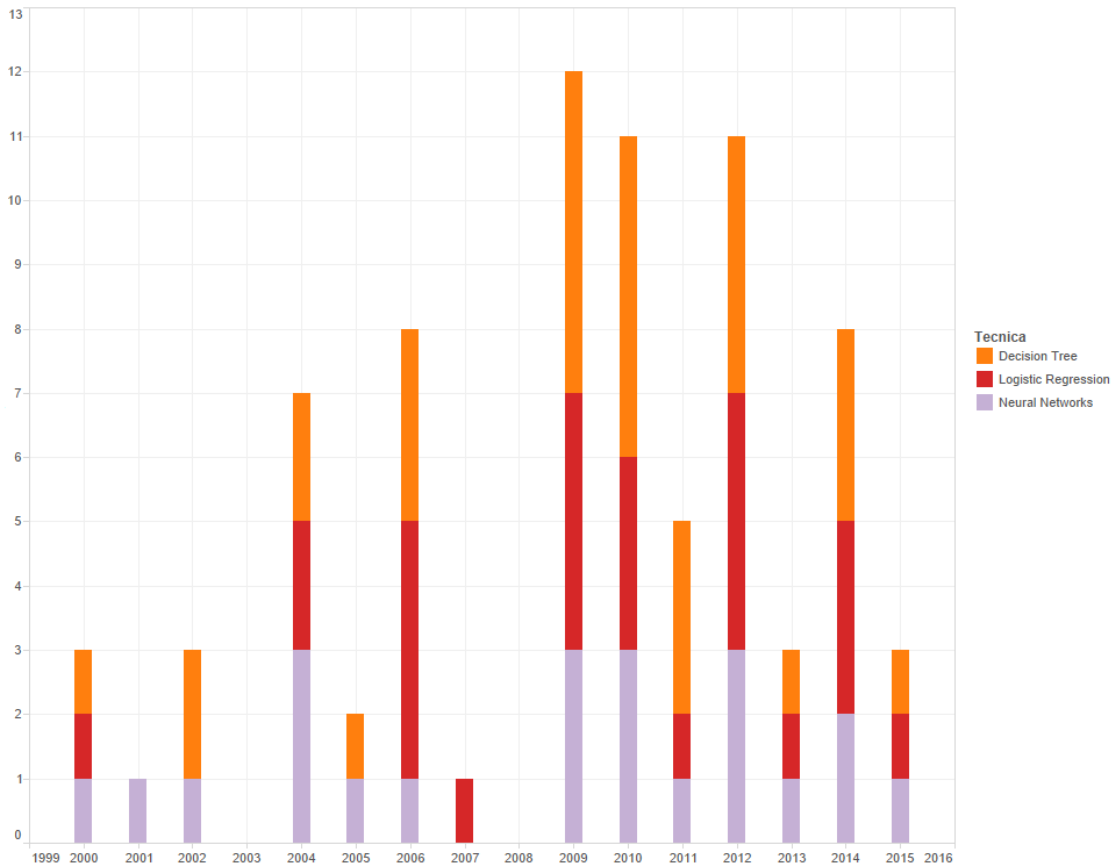


Figura 17: Barras: Classificação por Ano & Técnicas de Machine Learning com Filtro

4.7 Classificação por Área de Aplicação & Técnicas de Machine Learning.

Outras duas classificações foram combinadas aqui nesta seção, onde as figuras 18 e 19 mostram uma análise similar a que foi feita pelas figuras 16 e 17, sendo neste caso utilizado as classificações de Área de Aplicação e Técnicas de Machine Learning.

Logo, a figura 18, nos mostra a classificação combinada por Área de Aplicação e Técnicas de Machine Learning, para uma visão geral de todas as técnicas encontradas.

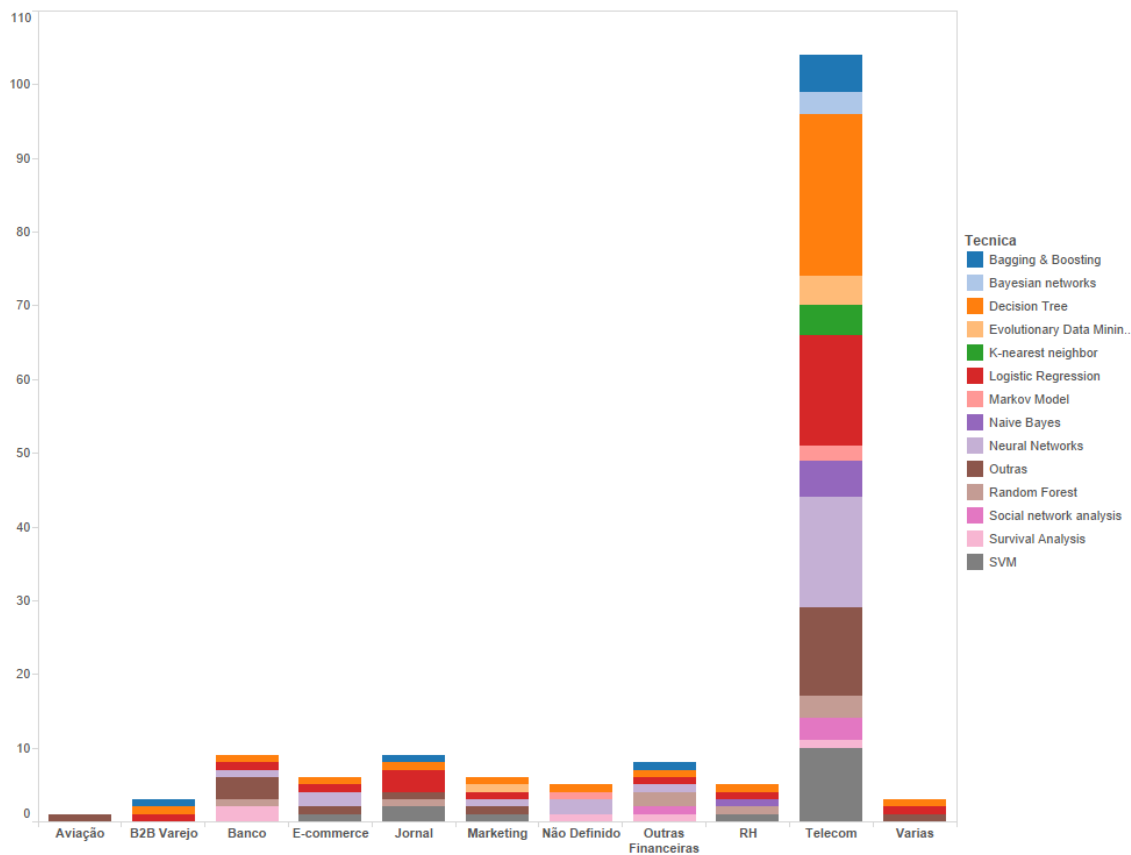


Figura 18: Barras: Classificação Área de Aplicação / Técnicas de Machine Learning.

E a figura 19, nos mostra que das 3 técnicas predominantes no estudo, pelo menos

duas delas aparecem em estudos de todas as diferentes áreas de aplicação encontradas na pesquisa. O que comprova a abrangência de aplicabilidade destas 3 técnicas, o que entendemos que colabora com o efeito predominante das mesmas.

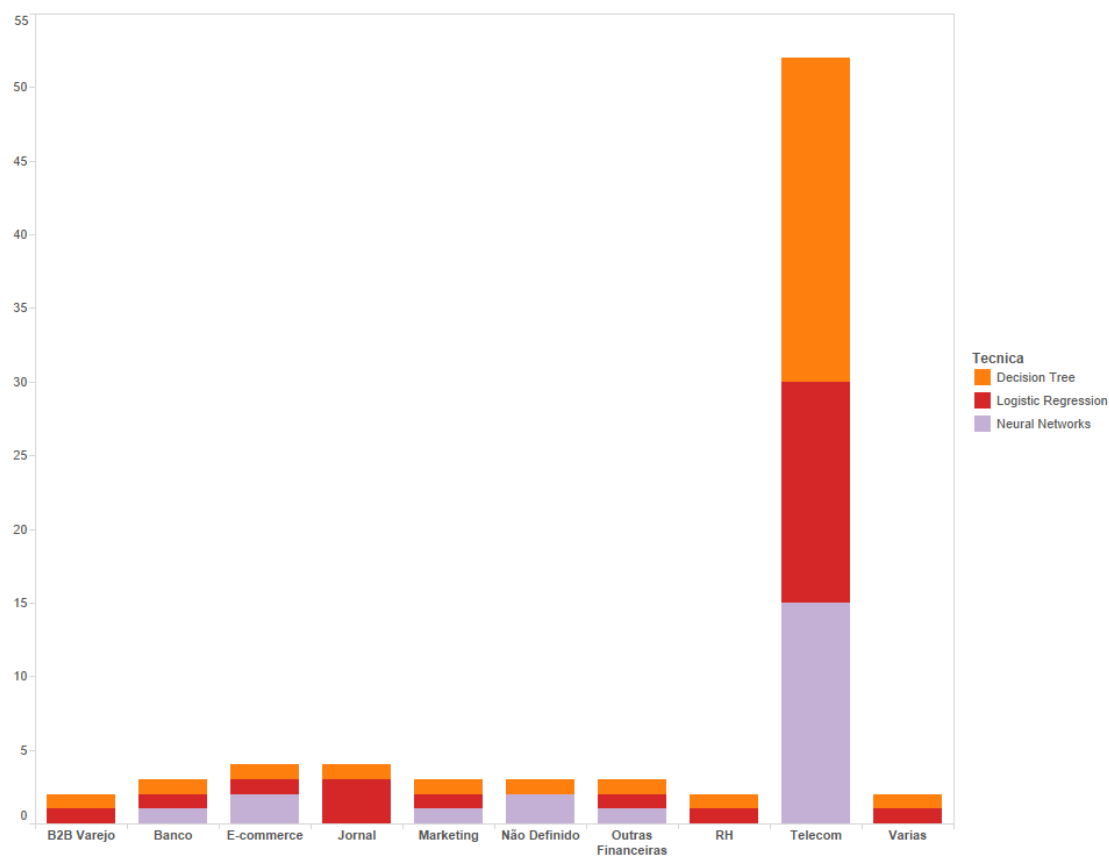


Figura 19: Barras: Classificação Área de Aplicação / Técnicas de Machine Learning.

4.8 Classificação por Citações

Nesta seção introduzimos os resultados obtidos através do estudo das citações de cada artigo utilizado, que por nossas pesquisas, provavelmente, é um estudo ainda inédito neste tipo de trabalho. Pretende-se com isso caracterizar o nível de influência de cada artigo, porém apenas limitado ao nível, sem extrair o(s) motivo(s) que está gerando tais níveis de influência. Para isso, com os dados dos artigos já preparados, como já explicado em seções anteriores, foram contabilizadas as quantidades de citações recebidas por cada um dos artigos e o valor atribuído a cada um dos artigos como mais um atributo na planilha de dados para análises. Além disto, foi construída uma rede, tecnicamente denominada como um grafo direcionado, onde os nós são os artigos e as arestas representam cada uma das citações dos 80 artigos reunidos no trabalho. Denominou-se esta rede de Rede de Influência, que foi construída através da ferramenta Gephi, com a qual novas métricas para a rede foram geradas e apresentadas aqui nesta seção do trabalho.



Figura 20: Classificação por Citações.

A figura 20 mostra a classificação dos artigos por suas Citações. Nesta figura estão representados todos os artigos citados. Pelo gráfico nota-se que 1610 diferentes artigos foram citados, isso de um total de 2660 citações. Nota-se também que a grande maioria, 76,34%(1229) dos artigos foram citados apenas uma vez, 12,36% (199) foram citados 2 vezes e assim por diante conforme o gráfico. Na figura 21 estão

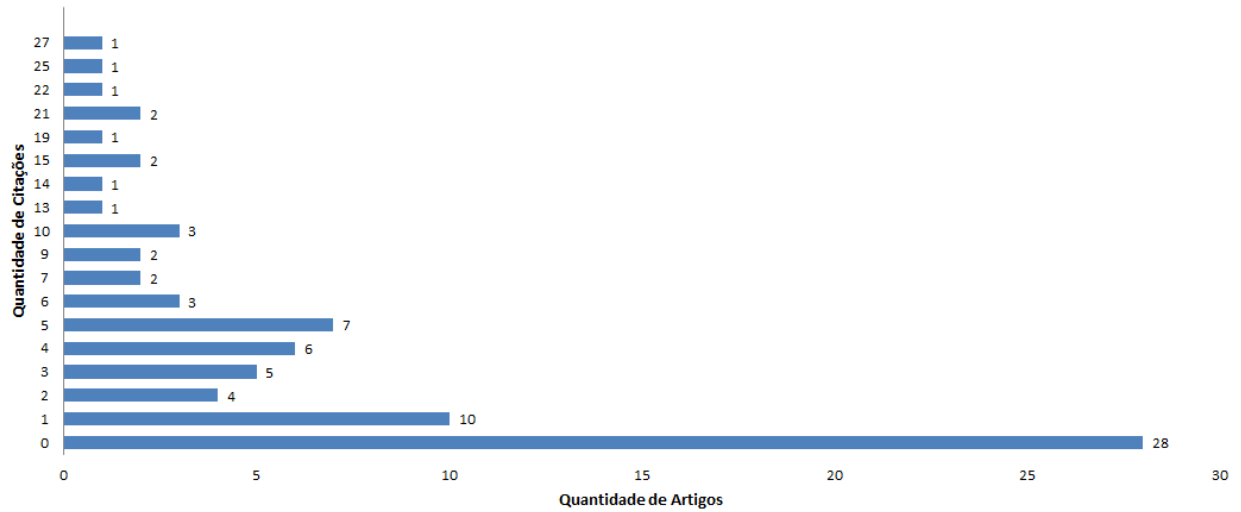


Figura 21: Classificação por Citações.

apenas as citações recebidas pelos 80 artigos utilizados na base do trabalho. Destes 80 artigos utilizados como base das classificações da pesquisa o artigo mais citado obteve 27 citações, que equivale a 34% dos artigos tendo-o citado, em seguida tem-se outro com 25 citações (31%), depois 22 (28%) e então 21(26%) e os demais abaixo de 20 citações recebidas. Ainda que 35% (28) dos artigos não tenham sido citados, outros, como foram vistos, possuem alto índice de citações. Talvez, mais uma vez isso possa ser justificado pela possibilidade de que haja uma tendência de que artigos publicados em um periódico cite outros artigos publicados no mesmo periódico.

Como constatação, gostaríamos de checar se este resultado, isto é, se a distribuição dos números de citações deste grupo de artigos segue, conforme existe na literatura acadêmica (Newman, 2005), uma distribuição de potência (do inglês, Power Law) ou também ainda conhecida como distribuição de Pareto ou Lei de Zipf. Para checar isso, ainda segundo Newman (2005), um dos métodos mais simples e utilizados, consiste em plotar um histograma em um gráfico com escalas *LogLog* e observar se desta forma obtém-se algo aproximado a uma reta. Portanto, como a discussão dos métodos de ajuste de uma lei de potência não fazem parte do objetivo principal deste trabalho, considerou-se válida essa técnica e seguindo o procedimento, conforme a figura 22 abaixo, constatamos o ajuste de nossos dados a esta teoria. Ainda foi traçada uma linha de regressão de potência, com uma medida de performance de $R^2 = 0,90$.

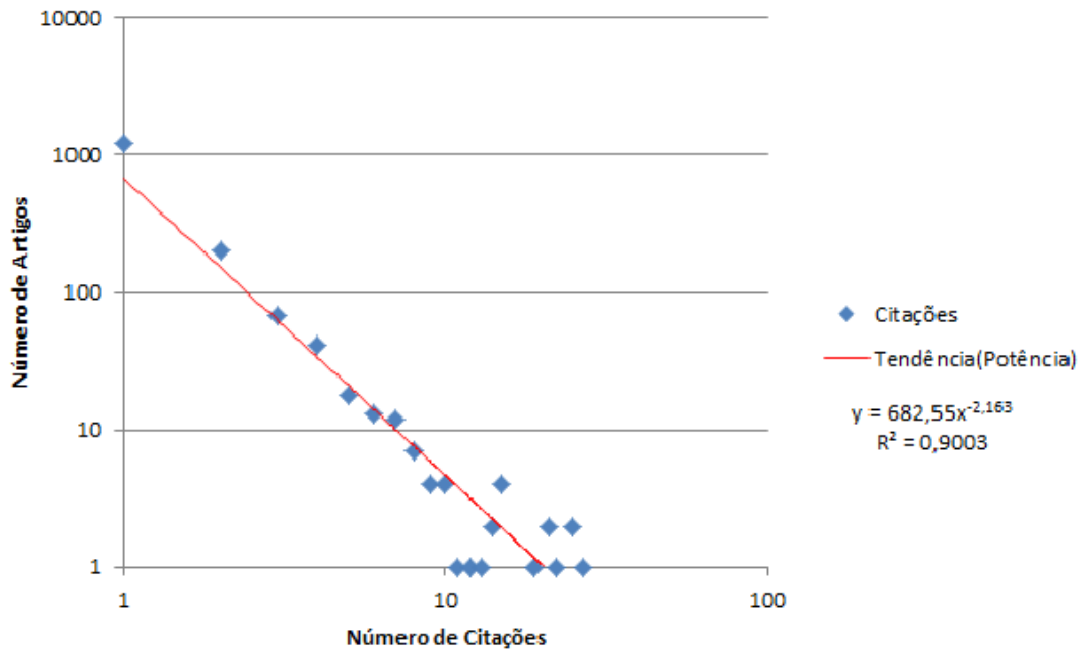


Figura 22: Distribuição das quantidades de Citações por modelagem de Potência.

Tomando-se a classificação dos artigos por números de citações, a qual, deliberadamente associou-se ao nível de influência dos artigos, a figura 23 mostra uma primeira representação e resultados dos níveis de influência dos artigos. A figura 23 se divide em três ilustrações. A primeira combina o valor do nível de influência com as Técnicas de Machine Learning. Isto é, representa os níveis de influência (ou quantidades de citações dos artigos), onde cada valor representa uma categoria posicionada no eixo horizontal. E para cada uma destas categorias de níveis, a mesma foi subdividida de acordo com as técnicas de Machine Learning encontradas em cada um destes níveis. A segunda ilustração, faz a mesma representação, porém filtrada, com o objetivo de destacar a partir do nível 10 de influência. Por último, tem-se o gráfico de classificação ponderada das técnicas de Machine Learning, isto é, cada técnica encontrada foi multiplicada pelo seu valor de influência e posteriormente somada ao total da sua categoria. Podemos comparar este gráfico com o da figura 10 onde percebemos que as 3 técnicas que mais apareceram no estudo de acordo com a figura 10, também prevalecem nos artigos com maior nível de influência. No entanto, houve uma aproximação ainda maior entre Regressão Logística e Redes Neurais, que



Figura 23: Classificação por Citações & Técnicas de Machine Learning.

por sua vez se distanciaram mais de Árvores de Decisão. Isso acontece, pois se olharmos no primeiro gráfico, percebemos que o artigo com maior nível de influência é composto apenas da técnica de Árvores de Decisão. A classificação das 4 primeiras técnicas, contudo, não se alterou. Porém, para as demais técnicas pode-se perceber alterações na ordem de classificação geral.

Conforme mencionado, foi então gerada a partir dos dados de citações dos artigos a Rede de Influência dos 80 artigos. A tabela a seguir, resume as características da rede gerada com auxílio da ferramenta Gephi, na qual se basearam todas as variações posteriores da rede. Vale ratificar que esta rede considera apenas as citações de seus 80 artigos, sem extensão às citações das citações e assim por diante. E que, o objetivo com isso neste trabalho, é produzir dados exploratórios e informativos, o que significa que não será aprofundado estudos sobre o posicionamento de cada artigo nas variações de redes geradas.

Tabela 3: Características da Rede de Influências

Parâmetro	Valor
Quantidade de nós gerados	1638
Quantidade de arestas geradas	2636
Diâmetro da rede	5
Modularidade da rede	0,635
Comprimento médio de caminho	2,518
Grafo Dirigido	SIM
Filtros ativos: Nós com # Citações	≥ 4
Quantidade de Comunidades Geradas	20
Total de Shortest Paths	24740

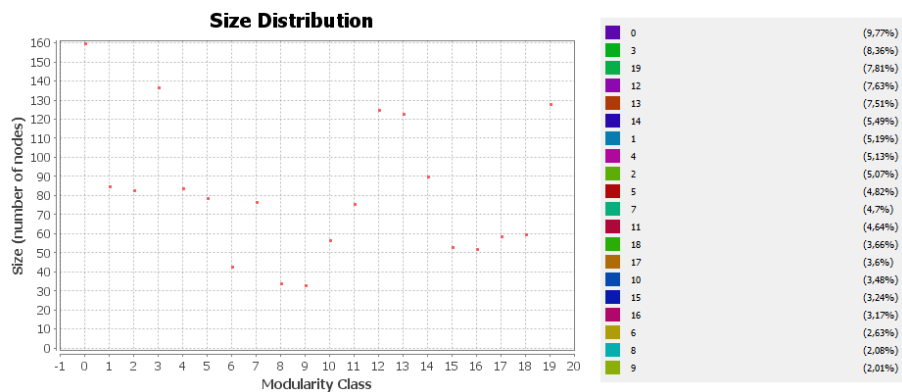


Figura 24: Comunidades detectadas e utilizadas pela ferramenta.

A figura 24 mostra a quantidade de comunidades criadas pela ferramenta, bem como a quantidade de artigos a elas associados. Para cada uma das comunidades, foi estabelecida uma cor, aleatoriamente, para então identificar cada nó da rede de influência de acordo com sua respectiva comunidade. Esta padronização foi utilizada em todas as seguintes redes apresentadas neste trabalho, para as quais foram sendo variadas as métricas de rede.

62

Tabela 4: Tabela ranking de Grau de Entrada

ALIAS	ID	ARTIGO	REF.	TÉCNICA	GRAU DE ENTRADA	CLASSE
C.P.Wei	80	Turning telecommunications call details to churn prediction- a data mining approach	Wei and Chin (2002)	Decision Tree(C4.5)	27	2
Hung	9	Applying data mining to telecom churn management	Hung, Yen and Wang (2006)	Decision Tree and NN	25	4
Buckinx	N/D	Customer base analysis: Partial defection of behaviorally-loyal clients in a non-contractual FMCG retail setting.	N/D	N/D	25	0
Coussement	6	Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques	Coussement and Van den Poel (2008)	SVM	22	4
Van den Poel	32	Customer attrition analysis for financial services using proportional hazard models	Van den Poel and Lariviere (2004)	Survival analysis	21	2
Mozer	8	Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry	Mozer et al. (2000)	Logistic Regression, Decision Trees, NN, and boosting	21	1
Au W., Chan C.C	21	A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction	Au, Chan and Yao (2003)	data mining by evolutionary learning (DMEL)	19	2
Lemmens	55	Bagging and Boosting Classification Trees to Predict Churn	Lemmens and Croux (2006)	Bagging and Boosting, Decision Trees	15	2
Neslin2	25	Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models	Neslin et al. (2006)	Logistic regression, decision trees	15	1
Quinlan	N/D	C4.5: Programs for Machine Learning	N/D	N/D	15	0

A figura 26 mostra a rede de influências caracterizada pelo PageRank de seus nós. Isto é, foi atribuído a cada nó da rede um tamanho proporcional ao valor de seu PageRank. O PageRank é um algoritmo criado pelo Google e mede a importância de uma página, no caso da internet, contabilizando a quantidade e qualidade de links apontando para ela. Não é o único algoritmo utilizado pelo Google para classificar páginas da internet, mas é o primeiro utilizado pela companhia e o mais conhecido.

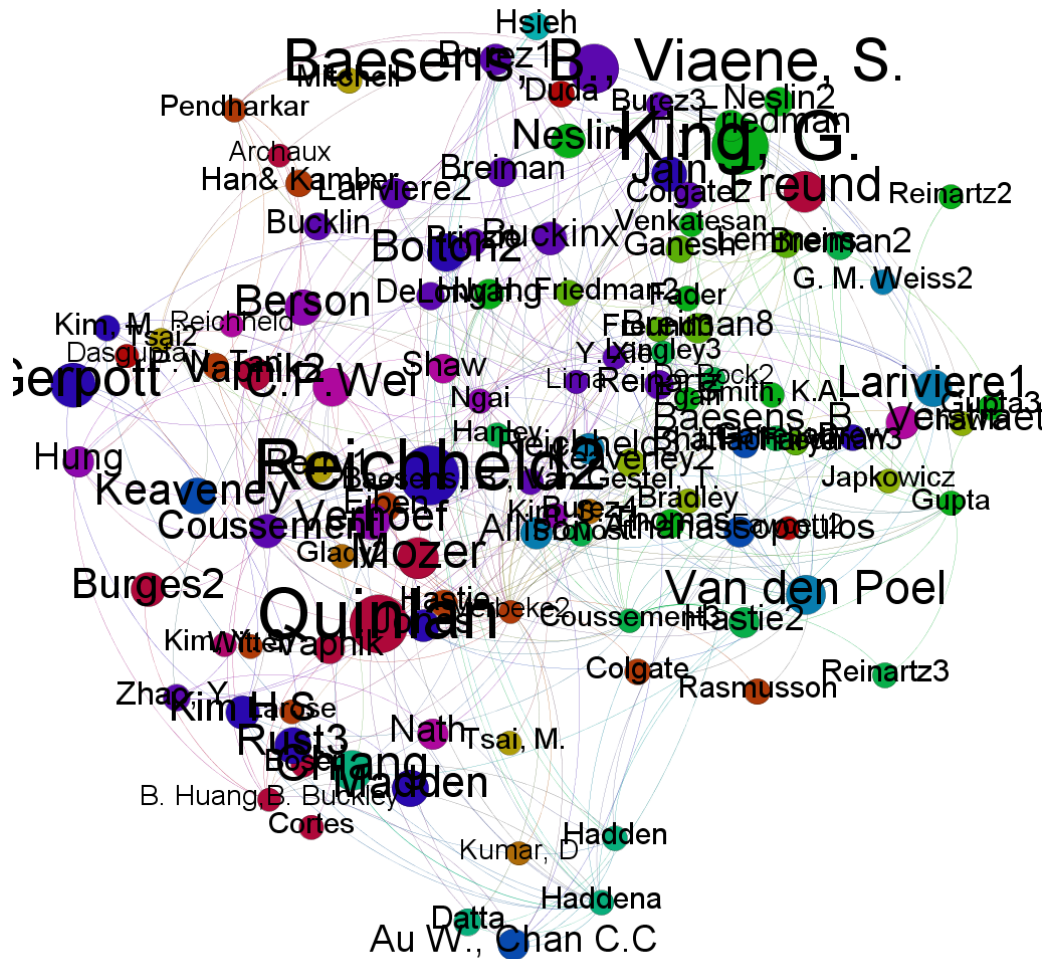


Figura 26: Rede de Influências, Métrica: PageRank.

Tabela 5: Tabela ranking (top10) por PageRank

ALIAS	ID	ARTIGO	REF.	TÉCNICA	PAGERANK	CLASSE
Reichheld2	N/D	The loyalty effect: The hidden force behind growth,profit and lasting value.	N/D	N/D	0,0193	14
Quinlan	N/D	C4.5: Programs for Machine Learning	N/D	N/D	0,0189	11
King, G.	N/D	C4.5: Programs for Machine Learning	N/D	N/D	0,0183	3
Bacens, B., Vi-aene, S.	N/D	Bayesian neural network learning for repeat purchase modeling in direct marketing.	N/D	N/D	0,0158	0
Gerpott	53	Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market	Gerpott, Rams and Schindler (2001)	LISREL	0,0141	14
Mozer	8	Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry	Mozer et al. (2000)	Logistic Regression, Decision Trees, NN, and boosting	0,0129	11
Freund	N/D	Experiments with a new boosting algorithm.	N/D	N/D	0,0129	11
Chiang	26	Goal-oriented sequential pattern for network banking churn analysis	Chiang et al. (2003)	Sequential analysis	0,0124	7
Van den Poel	32	Customer attrition analysis for financial services using proportional hazard models	Van den Poel and Lariviere (2004)	proportional hazard model (survival analysis)	0,0123	1
C.P.Wei	80	Turning telecommunications call details to churn prediction- a data mining approach	Wei and Chiu (2002)	Decision Tree(C4.5)	0,0119	4

A figura 27 mostra a rede de influências caracterizada pela Centralidade de Intermediação de seus nós. Isto é, foi atribuído a cada nó da rede um tamanho proporcional ao valor de sua Centralidade de Intermediação. Centralidade de Intermediação mede a frequência com que um nó aparece nos caminhos mais curtos entre nós da rede. Em geral essa medida reflete a quantidade de controle que cada nó exerce na interação com os demais nós da rede. Em outras palavras, um nó com alta Centralidade de Intermediação tem grande influência no fluxo da rede, assumindo que em geral o fluxo de rede busca os caminhos mais curtos.

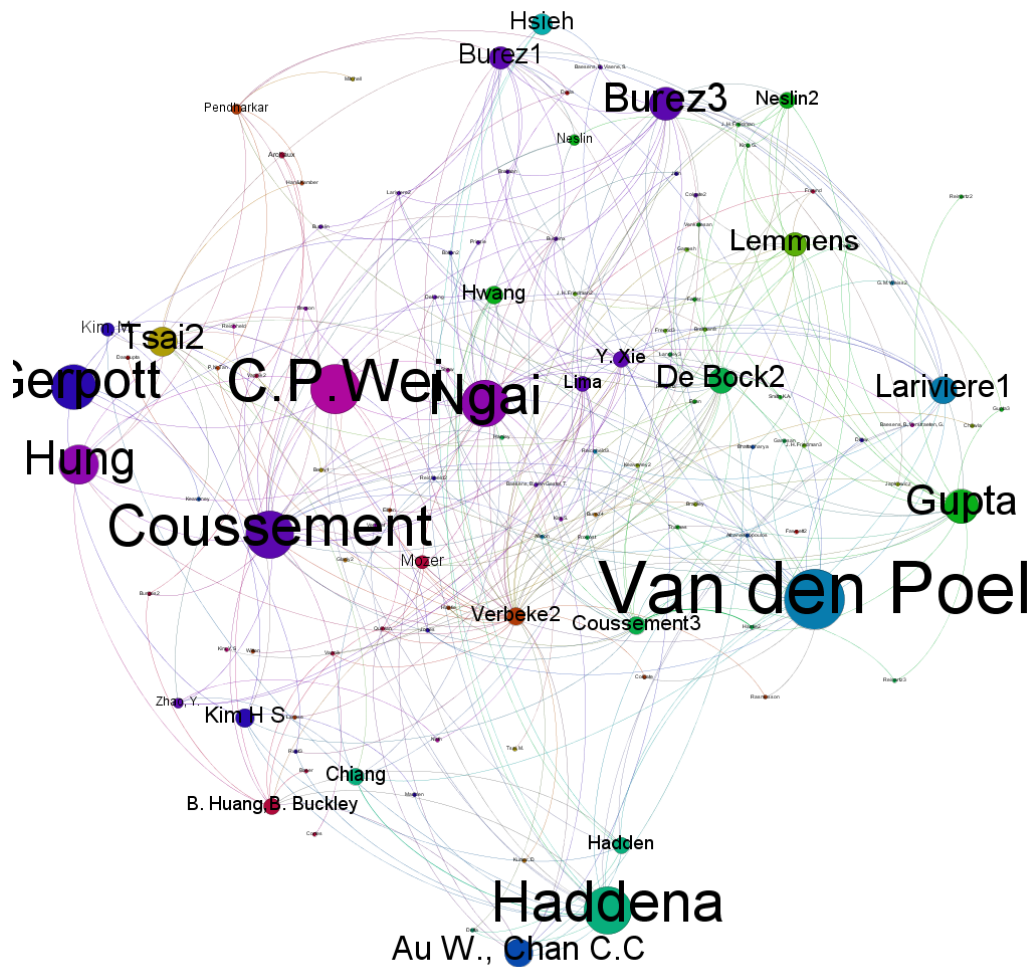


Figura 27: Rede de Influências, Métrica: Centralidade de Intermediação.

Tabela 6: Tabela ranking (top10) por Centralidade de Intermediação

ALIAS	ID	ARTIGO	REF	TÉCNICA	INTERMEDIÇÃO	CLASSE
Van den Poel	32	Customer attrition analysis for financial services using proportional hazard models	Van den Poel and Lariviere (2004)	proportional hazard model (survival analysis)	2790,72	1
C.P.Wei	80	Turning telecommunications call details to churn prediction- a data mining approach	Wei and Chiu (2002)	Decision Tree(C4.5)	2267,64	4
Coussement	6	Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques	Coussement and Van den Poel (2008)	SVM	2174,44	0
Haddena	31	Computer Assisted Customer Churn Management: State-Of-The-Art and Future Trends	Hadden et al. (2007)	NA	2167,20	7
Ngai	46	Application of data mining techniques in customer relationship management: A literature review and classification	Ngai, Xiu and Chau (2009)	NA	2129,67	12
Gerpott	53	Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market	Gerpott, Rams and Schindler (2001)	LISREL	2026,65	14
Hung	9	Applying data mining to telecom churn management	Hung, Yen and Wang (2006)	Decision Tree and NN	1757,11	12
Gupta	56	Modeling Customer Lifetime Value	Gupta et al. (2006)	Hazard Model,Markov Model	1502,42	3
Burez3	38	Handling class imbalance in customer churn prediction	Burez and Van den Poel (2009)	Logistic Regression and Decision Trees	1447,23	0
Tsai2	34	Customer churn prediction by hybrid neural networks	Tsai and Lu (2009)	NN	1250,15	6

A figura 28 mostra a rede de influências caracterizada pela Centralidade de Proximidade de seus nós. Isto é, foi atribuído a cada nó da rede um tamanho proporcional ao valor de sua Centralidade de Proximidade. Centralidade de Proximidade mede a distância média de um determinado nó inicial para todos os demais nós da rede. Com isso, quanto mais central for nó, menor será o somatório total das distâncias para os demais nós da rede. Contudo, pela formulação utilizada para esta métrica, ainda assim os nós mais centrais são representados pelas medidas mais altas. A Centralidade de Proximidade também pode ser interpretada pela velocidade com que uma informação flui de um determinado nó da rede para os demais nós atingíveis através deste.

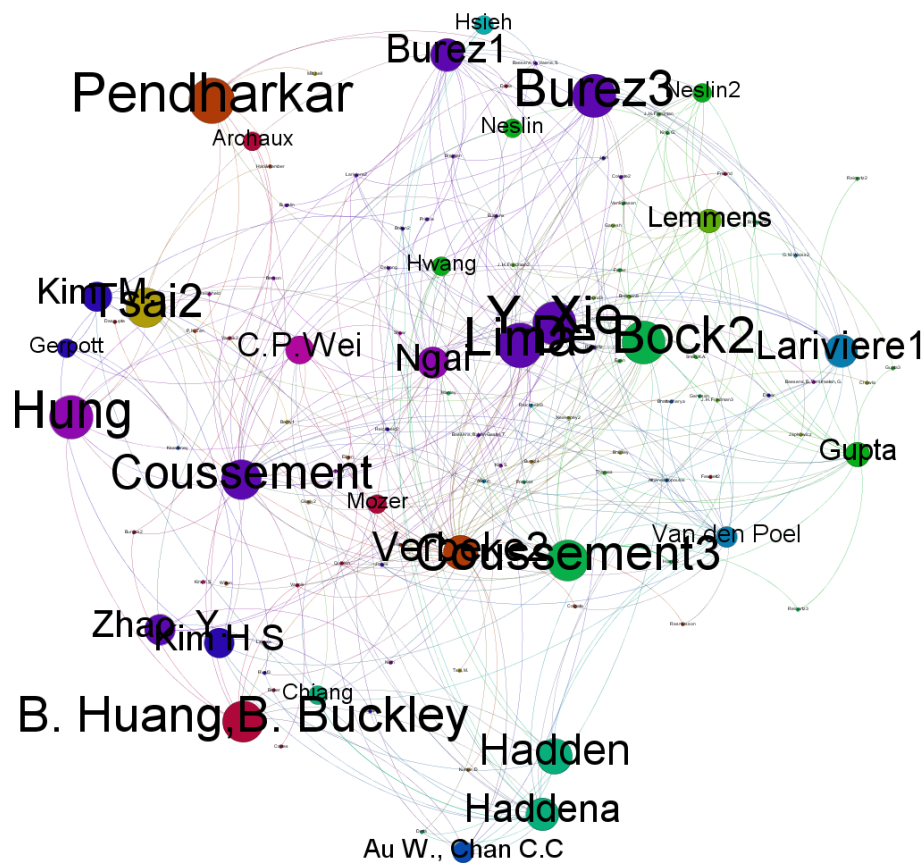


Figura 28: Rede de Influências, Métrica: Centralidade de Proximidade.

Tabela 7: Tabela ranking (top10) por Centralidade de Proximidade

ALIAS	ID	ARTIGO	REF	TÉCNICA	PROXIMIDADE	CLASSE
Castanedo	1	Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network	Castanedo et al. (2014)	Deep Learning	3,27	13
Ballings	67	Customer event history for churn prediction: How long is long enough?	Ballings and Van den Poel (2012)	logistic regression, classification trees, bagging	3,11	19
Benoit2	75	Improving customer retention in financial services using kinship network information	Benoit Van den Poel (2012)	Social network analysis(SNA), Random forests	3,02	0
Olle	7	A Hybrid Churn Prediction Model in Mobile Telecommunication Industry	Olle and Cai (2014)	Logistic Regression	2,98	4
K. Kim	72	Improved churn prediction in telecommunication industry by analyzing a large network	Kim, Jun and Lee (2014)	Network analysis, Community detection, Logistic Regression, NN	2,94	18
Vafeiadis	57	A comparison of machine learning techniques for customer churn prediction	Vafeiadis et al. (2015)	NN,SVM, Decision Trees, Logistic Regression, Boosting, Naive Bayes	2,87	4
De Bock5	78	Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models	De Bock and Van den Poel (2012)	Ensemble classification, Generalized additive models (GAMs)	2,87	19
Kisioglu	62	Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey	Kisioglu and Topcu (2011)	Bayesian Belief Networks	2,85	4
Idris2	64	Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies	Idris, Rizwan and Khan (2012)	Random Forest (RF) and K Nearest Neighbour (KNN)	2,79	9
Keramatia	73	Improved churn prediction in telecommunication industry using Data Mining techniques	Keramati et al. (2014)	ANN,KNN,SVM,Decision tree	2,75	5

A figura 29 mostra a rede de influências caracterizada pela Excentricidade de seus nós. Isto é, foi atribuído a cada nó da rede um tamanho proporcional ao valor de sua Excentricidade. Excentricidade mede a distância de um determinado nó inicial até o nó mais distante dele na rede. Note que a máxima Excentricidade de uma rede é dada pelo seu diâmetro, que como vimos, neste caso é igual a 5.

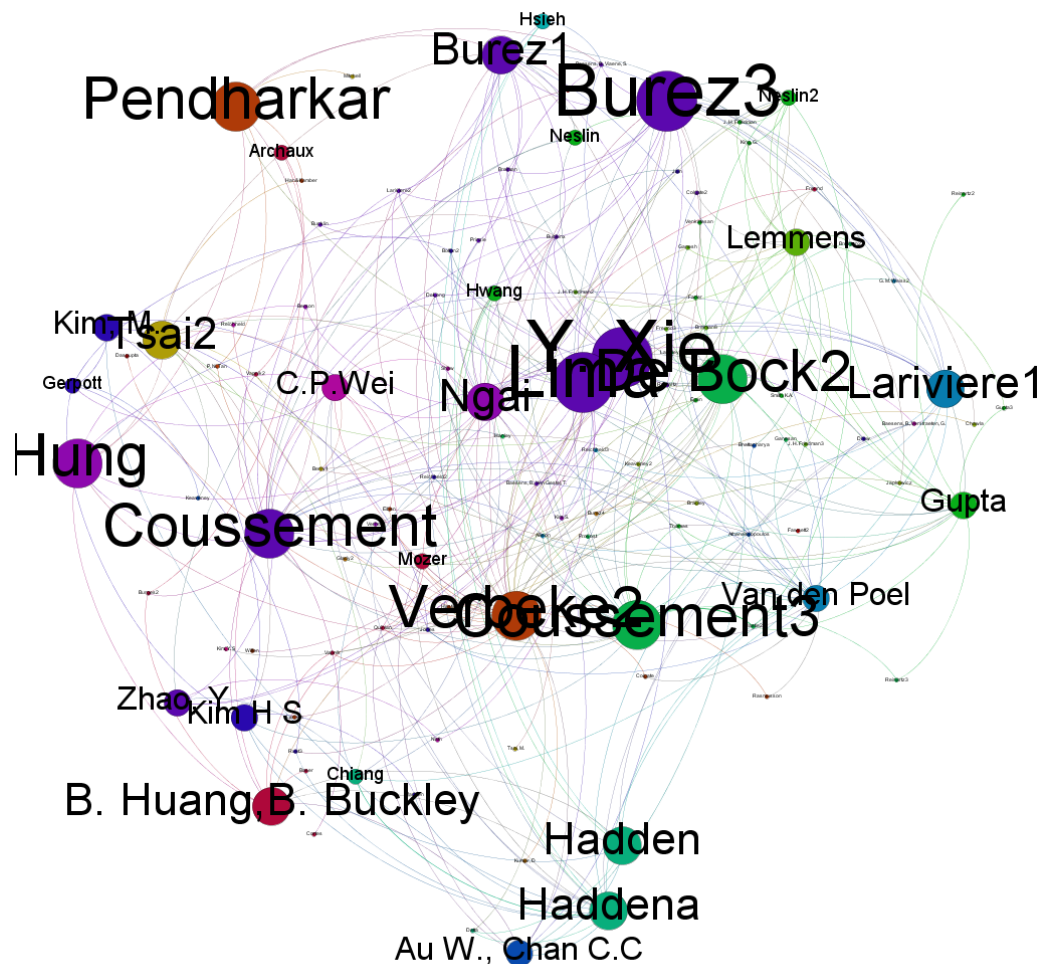


Figura 29: Rede de Influências, Métrica: Excentricidade.

Tabela 8: Tabela ranking (top10) por Excentricidade

ALIAS	ID	ARTIGO	REF	TÉCNICA	EXCENTRICIDADE	CLASSE
Ballings	67	Customer event history for churn prediction: How long is long enough?	Ballings and Van den Poel (2012)	logistic regression, classification trees, bagging	5	19
Benoit2	75	Improving customer retention in financial services using kinship network information	Benoit and Van den Poel (2012)	Social network analysis (SNA), Random forests	5	0
De Bock5	78	Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models	De Bock and Van den Poel (2012)	Ensemble classification Generalized additive models (GAMs)	5	19
Lima	50	Domain Knowledge Integration in Data Mining using Decision Tables: Case Studies in Churn Prediction	Lima, Mues and Baesens (2009)	Logistic Regression and Decision Trees	5	0
Burez3	38	Handling class imbalance in customer churn prediction	Burez and Van den Poel (2009)	Logistic Regression and Decision Trees	5	0
Y. Xie	2	Customer churn prediction using improved balanced random forests	Xie et al. (2009)	Random Forests	5	0
Castanedo	1	Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network	Castanedo et al. (2014)	Deep Learning	4	13
Olle	7	A Hybrid Churn Prediction Model in Mobile Telecommunication Industry	Olle and Cai (2014)	Logistic Regression	4	4
K. Kim	72	Improved churn prediction in telecommunication industry by analyzing a large network	Kim, Jun and Lee (2014)	Network analysis, Community detection, Logistic Regression, NN	4	18
Vafeiadis	57	A comparison of machine learning techniques for customer churn prediction	Vafeiadis et al. (2015)	NN, SVM, Decision Trees, Logistic Regression, Boosting, Naive Bayes	4	4

5 Conclusões da Pesquisa, Limitações e Trabalhos Futuros.

Este trabalho destina-se a fazer uma síntese de pesquisa na aplicação, em especial, de técnicas de Machine Learning, para a análise preditiva de Churn. Embora esta revisão não tenha a pretensão de ser exaustiva, ela traz ensejos razoáveis bem como mostra a incidência de pesquisa sobre este tema.

Constatamos a grande incidência de estudos de Churn bem como uma breve percepção de estudos adjacentes focados em retenção do cliente, que mostram que hoje em dia as empresas que têm seu negócio sustentado por uma base de assinantes, têm tido grande preocupação com as questões de lealdade do cliente e consequente perda de receita em função de perdas de clientes. E uma das maneiras mais eficientes de minimizar este problema é através de modelos preditivos de Churn, para os quais as técnicas de Machine Learning em especial, ajudam na identificação destes clientes para que os investimentos e ações de retenção sejam direcionados da forma mais eficaz possível.

Os resultados apresentaram claramente 3 técnicas em destaque nesta pesquisa: Árvores de Decisão, Regressão Logística e Redes Neurais, embora várias outras técnicas tenham se manifestado. Pudemos perceber também a importância da utilização de mais de uma técnica simultaneamente para comparações e busca de melhores resultados.

Ainda do ponto de vista de resultados, foram encontrados artigos publicados entre os anos 2000 e 2015, sendo observada uma maior concentração desta área de pesquisa nos anos mais recentes. Percebemos também a maior concentração deste tipo de estudo na área de Telecomunicações. E que de forma não esperada obtivemos grandes concentrações nas fontes de origem dos artigos tanto do ponto de vista de Repositórios quanto de Periódicos de publicação. E por último, mas não menos importante, construiu-se uma rede de influências entre os artigos observados, podendo-se perceber através de diferentes métricas a forma com que este grupo de artigos exercia influência entre si.

Podemos, por outro lado, identificar algumas possíveis limitações deste estudo, melhorias e futuras pesquisas.

Foram observados relatos ao longo desta pesquisa, que não estão diretamente ligados a limitações da mesma, mas de problemas com os quais os pesquisadores se depararam em seus estudos de análise preditiva de Churn. Alguns relatam conjuntos de dados incompletos ou faltantes (Huang et al., 2010). Outros, o oposto, isto é, quantidades imensas de dados de difícil estruturação e manipulação (Lin, Tzeng and Chin, 2011). O que de fato, mais uma vez, nos leva ao entendimento da grande

utilização das técnicas de Árvores de Decisão que se apresentam de maneira mais robusta a estes problemas. Ou ainda, na opinião de [Xie et al. \(2009\)](#); [Huang et al. \(2010\)](#), o maior desafio deste estudo é a questão de desbalanceamento dos dados, que se refere ao fato da taxa dos potenciais clientes classificados como Churners em relação aos clientes regulares ser muito baixa. Tomando-se alguns exemplos, a taxa de clientes classificados com potencial de Churn está em torno de valores de 2% ou 3% do total da base, o que certamente é um problema para os modelos preditivos.

Prazo e recursos utilizados também são vistos como um fator limitante. Por se tratar de um trabalho de conclusão de curso de mestrado, havia um tempo limitado para completar o estudo, além da obrigatoriedade da pesquisa ser individual, com sua devida orientação, porém executada individualmente.

No que diz respeito ao prazo, foi despendido um esforço e tempo muito grandes na elaboração da rede de citações. Como explicado anteriormente se tratou de um trabalho parcialmente manual com poucos recursos automatizados. Desta forma o cuidado foi redobrado para evitar erros tomando-se assim uma grande porção de tempo total da elaboração do corpus de documento do trabalho. Visamos que em trabalhos futuros, este método possa ser substituído por alguma ferramenta existente ou possa ser desenvolvida uma. Que ora não foi desenvolvida, por falta de tempo ou por entendermos que isso nos distanciaria do foco deste trabalho devido a complexidade desta tarefa que poderia se transformar em outra dissertação de mestrado no âmbito de sistemas de recuperação de informação.

Da mesma forma, por limitações de tempo e foco do trabalho, não foi feito um aprofundamento na teoria de redes ou teoria de grafos, que talvez, ajudaria a completar a interpretação e comentários das redes de citações criadas pela ferramenta de software utilizada.

Provavelmente mais artigos poderiam se juntar aos artigos utilizados, adotando novos critérios de pesquisa, como modificando os termos utilizados na busca, uma vez que eventuais artigos que tratassem do mesmo assunto porem utilizando termos diferentes ou omitindo-os não foram considerados. Outra questão é a língua utilizada nos artigos, que pelos termos utilizados na busca, esta foi direcionada diretamente para a língua inglesa e assim as demais línguas não foram consideradas na pesquisa. Novas ferramentas de busca talvez pudessem ser utilizadas assim como novas estratégias de buscas, como um exemplo, focadas de forma mais ampla e minuciosa por repositórios ou periódicos a fim de confirmar ou tentar ainda mais evitar concentrações inesperadas como as que ocorreram.

Embora entenda que com a melhoria destas limitações supracitadas poderíamos ter resultados mais amplos e talvez mais precisos, acreditamos como mencionado, que nosso resultado fornece uma análise significativa e com informações relevantes

sobre o assunto.

Para trabalhos futuros, além destas melhorias, pensa-se na possibilidade de acrescentar novas dimensões de classificações tais como: Desempenho dos resultados obtidos com cada técnica; Utilização de técnicas de pré-processamento da base de dados, que é uma dimensão importante para se entender como os pesquisadores estão lidando com o problema de desbalanceamento dos dados, que é inerente ao estudo do Churn; E Dimensões e tipos de base de dados utilizada. No entanto, enxergamos uma maior dificuldade nestas dimensões por se tratarem de informações menos homogêneas para tabular, e algumas vezes, como no detalhamento da base de dados, por exemplo, se tratarem de informações muitas vezes confidenciais e não fornecidas.

Anseia-se também, como continuação do trabalho, a aquisição de uma base de dados e processamento das mesmas de acordo com as principais técnicas obtidas no resultado desta pesquisa. E, desta forma, podermos efetivamente analisar e comparar as diferentes técnicas de Machine Learning diante de uma mesma base de dados.

Referências

- Aeron, Harsha, Ashwani Kumar and Moorthy Janakiraman. 2010. “Application of data mining techniques for customer lifetime value parameters: a review.” *International Journal of Business Information Systems* 6(4):514–529.
- Ahn, Jae-Hyeon, Sang-Pil Han and Yung-Seop Lee. 2006. “Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry.” *Telecommunications policy* 30(10):552–568.
- Ali, Özden Gür and Umut Arıtürk. 2014. “Dynamic churn prediction framework with more effective use of rare event data: The case of private banking.” *Expert Systems with Applications* 41(17):7889–7903.
- Archaux, Cedric, Arnaud Martin and Ali Khenchaf. 2004. An SVM based churn detector in prepaid mobile telephony. In *Information and Communication Technologies: From Theory to Applications, 2004. Proceedings. 2004 International Conference on*. IEEE pp. 459–460.
- Au, Wai-Ho, Keith CC Chan and Xin Yao. 2003. “A novel evolutionary data mining algorithm with applications to churn prediction.” *Evolutionary Computation, IEEE Transactions on* 7(6):532–545.
- Ballings, Michel and Dirk Van den Poel. 2012. “Customer event history for churn prediction: How long is long enough?” *Expert Systems with Applications* 39(18):13517–13522.
- Benoit, Dries F and Dirk Van den Poel. 2012. “Improving customer retention in financial services using kinship network information.” *Expert Systems with Applications* 39(13):11435–11442.
- Berson, A, S Smith and K Thearling. 2000. *Building data mining applications for CRM*. McGraw-Hill.
- Burez, Jonathan and Dirk Van den Poel. 2007. “CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services.” *Expert Systems with Applications* 32(2):277–288.
- Burez, Jonathan and Dirk Van den Poel. 2009. “Handling class imbalance in customer churn prediction.” *Expert Systems with Applications* 36(3):4626–4636.

- Castanedo, Federico, Gabriel Valverde, Jaime Zaratiegui and Alfonso Alfonso Vazquez. 2014. "Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network."
- Chiang, Ding-An, Yi-Fan Wang, Shao-Lun Lee and Cheng-Jung Lin. 2003. "Goal-oriented sequential pattern for network banking churn analysis." *Expert Systems with Applications* 25(3):293–302.
- Coussement, Kristof and Dirk Van den Poel. 2008. "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques." *Expert systems with applications* 34(1):313–327.
- Coussement, Kristof and Dirk Van den Poel. 2009. "Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers." *Expert Systems with Applications* 36(3):6127–6134.
- Coussement, Kristof, Dries F Benoit and Dirk Van den Poel. 2010. "Improved marketing decision making in a customer churn prediction context using generalized additive models." *Expert Systems with Applications* 37(3):2132–2143.
- De Bock, Koen W and Dirk Van den Poel. 2011. "An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction." *Expert Systems with Applications* 38(10):12293–12301.
- De Bock, Koen W and Dirk Van den Poel. 2012. "Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models." *Expert Systems with Applications* 39(8):6816–6826.
- Deri, Joya A and José MF Moura. 2014. Churn detection in large user networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE pp. 1090–1094.
- Dierkes, Torsten, Martin Bichler and Ramayya Krishnan. 2011. "Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks." *Decision Support Systems* 51(3):361–371.
- Drew, James H, DR Mani, Andrew L Betz and Piew Datta. 2001. "Targeting customers with statistical and data-mining techniques." *Journal of Service Research* 3(3):205–219.

- Eshghi, Abdolreza, Dominique Haughton and Heikki Topi. 2007. "Determinants of customer loyalty in the wireless telecommunications industry." *Telecommunications policy* 31(2):93–106.
- Gerpott, Torsten J, Wolfgang Rams and Andreas Schindler. 2001. "Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market." *Telecommunications policy* 25(4):249–269.
- Glady, Nicolas, Bart Baesens and Christophe Croux. 2009. "Modeling churn using customer lifetime value." *European Journal of Operational Research* 197(1):402–411.
- Gupta, Sunil, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V Kumar, Nathaniel Lin, Nalini Ravishanker and S Sriram. 2006. "Modeling customer lifetime value." *Journal of service research* 9(2):139–155.
- Hadden, John, Ashutosh Tiwari, Rajkumar Roy and Dymitr Ruta. 2007. "Computer assisted customer churn management: State-of-the-art and future trends." *Computers & Operations Research* 34(10):2902–2917.
- Hadden, John, Ashutosh Tiwari, Rajkumar Roy and Dymtr Ruta. 2006. Churn prediction using complaints data. In *Proceedings Of World Academy Of Science, Engineering and Technology*. Citeseer.
- Han, Jiawei and Micheline Kamber. 2006. *Data Mining: Concepts and Techniques*. Morgan Kaufmann - Elsevier.
- Hashmi, Nabgha, Naveed Anwer Butt and Dr Muddesar Iqbal. 2013. "Customer Churn Prediction in Telecommunication A Decade Review and Classification." *International Journal of Computer Science Issues* 10:271–282.
- Hsieh, Nan-Chen. 2004. "An integrated data mining and behavioral scoring model for analyzing bank customers." *Expert systems with applications* 27(4):623–633.
- Huang, Bing Quan, T-M Kechadi, Brian Buckley, G Kiernan, E Keogh and Tarrif Rashid. 2010. "A new feature set with new window techniques for customer churn prediction in land-line telecommunications." *Expert Systems with Applications* 37(5):3657–3665.
- Huang, Bingquan, Brian Buckley and T-M Kechadi. 2010. "Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications." *Expert Systems with Applications* 37(5):3638–3646.

- Huang, Bingquan, Mohand Tahar Kechadi and Brian Buckley. 2012. "Customer churn prediction in telecommunications." *Expert Systems with Applications* 39(1):1414–1425.
- Huang, Ying and Tahar Kechadi. 2013. "An effective hybrid learning system for telecommunication churn prediction." *Expert Systems with Applications* 40(14):5635–5647.
- Hung, Shin-Yuan, David C Yen and Hsiu-Yu Wang. 2006. "Applying data mining to telecom churn management." *Expert Systems with Applications* 31(3):515–524.
- Hwang, Hyunseok, Taesoo Jung and Euiho Suh. 2004. "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry." *Expert systems with applications* 26(2):181–188.
- Idris, Adnan, Muhammad Rizwan and Asifullah Khan. 2012. "Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies." *Computers & Electrical Engineering* 38(6):1808–1819.
- Institute, SAS. 2000. "Best Price in Churn Prediction." SAS Institute White Paper.
- Jahromi, Ali Tamaddoni, Stanislav Stakhovych and Michael Ewing. 2014. "Managing B2B customer churn, retention and profitability." *Industrial Marketing Management* 43(7):1258–1268.
- Karahoca, Adem and Dilek Karahoca. 2011. "GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system." *Expert Systems with Applications* 38(3):1814–1822.
- Keramati, Abbas, Ruholla Jafari-Marandi, Mohammed Aliannejadi, Iman Ahmadian, Mahdiah Mozaffari and Uldoz Abbasi. 2014. "Improved churn prediction in telecommunication industry using data mining techniques." *Applied Soft Computing* 24:994–1012.
- Kim, Hee-Su and Choong-Han Yoon. 2004. "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market." *Telecommunications policy* 28(9):751–765.
- Kim, Kyoungok, Chi-Hyuk Jun and Jaewook Lee. 2014. "Improved churn prediction in telecommunication industry by analyzing a large network." *Expert Systems with Applications* 41(15):6575–6584.

- Kim, Kyoungok and Jaewook Lee. 2012. "Sequential manifold learning for efficient churn prediction." *Expert Systems with Applications* 39(18):13328–13337.
- Kim, Moon-Koo, Myeong-Cheol Park and Dong-Heon Jeong. 2004. "The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services." *Telecommunications policy* 28(2):145–159.
- Kim, Yong Seog, Hyeseon Lee and John D Johnson. 2013. "Churn management optimization with controllable marketing variables and associated management costs." *Expert Systems with Applications* 40(6):2198–2207.
- Kincaid, J. W. 2003. *Customer relationship management: Getting it right*. Prentice Hall PTR.
- Kirui, Clement, Li Hong, Wilson Cheruiyot and Hillary Kirui. 2013. "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining." *IJCSI International Journal of Computer Science Issues* 10(2):1694–0784.
- Kisioglu, Pinar and Y Ilker Topcu. 2011. "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey." *Expert Systems with Applications* 38(6):7151–7157.
- Larivière, Bart and Dirk Van den Poel. 2004. "Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services." *Expert Systems with Applications* 27(2):277–285.
- Larivière, Bart and Dirk Van den Poel. 2005. "Predicting customer retention and profitability by using random forests and regression forests techniques." *Expert Systems with Applications* 29(2):472–484.
- Lejeune, M. 2001. "Measuring the impact of data mining on churn management." *Internet Research: Electronic Network Applications and Policy* 11(5):375–387.
- Lemmens, Aurélie and Christophe Croux. 2006. "Bagging and boosting classification trees to predict churn." *Journal of Marketing Research* 43(2):276–286.
- Lima, Elen, Christophe Mues and Bart Baesens. 2009. "Domain knowledge integration in data mining using decision tables: Case studies in churn prediction." *Journal of the Operational Research Society* 60(8):1096–1106.
- Lin, Chiun-Sin, Gwo-Hshiung Tzeng and Yang-Chieh Chin. 2011. "Combined rough set theory and flow network graph to predict customer churn in credit card accounts." *Expert Systems with Applications* 38(1):8–15.

- Liou, James JH. 2009. "A novel decision rules approach for customer relationship management of the airline market." *Expert systems with Applications* 36(3):4374–4381.
- Maldonado, Sebastián, Álvaro Flores, Thomas Verbraken, Bart Baesens and Richard Weber. 2015. "Profit-based feature selection using support vector machines—General framework and an application for customer retention." *Applied Soft Computing* 35:740–748.
- Miller, Hugh, Sandy Clarke, Stephen Lane, Andrew Lonie, David Lazaridis, Slave Petrovski and Owen Jones. 2009. "Predicting customer behaviour: The University of Melbourne's KDD Cup report." *The 2009 Knowledge Discovery in Data Competition (KDD Cup 2009) Challenges in Machine Learning, Volume 3* p. 43.
- Mozier, Michael C, Richard Wolniewicz, David B Grimes, Eric Johnson and Howard Kaushansky. 2000. "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry." *Neural Networks, IEEE Transactions on* 11(3):690–696.
- Mutanen, Teemu, Jussi Ahola and Sami Nousiainen. 2006. "Customer churn prediction—a case study in retail banking." *Practical Data Mining: Applications, Experiences and Challenges* p. 13.
- Neslin, Scott A, Sunil Gupta, Wagner Kamakura, Junxiang Lu and Charlotte H Mason. 2006. "Defection detection: Measuring and understanding the predictive accuracy of customer churn models." *Journal of marketing research* 43(2):204–211.
- Neslin, Scott, Sunil Gupta, Wagner Kamakura, Junxiang Lu and Charlotte Mason. 2004. "Defection detection: improving predictive accuracy of customer churn models." *Tuck School of Business, Dartmouth College*.
- Newman, M.E.J. 2005. "Power laws, Pareto distributions and Zipf's law." *Contemporary Physics* 46(5):323–351.
- Ngai, Eric WT. 2005. "Customer relationship management research (1992-2002) An academic literature review and classification." *Marketing intelligence & planning* 23(6):582–605.
- Ngai, Eric WT, Li Xiu and Dorothy CK Chau. 2009. "Application of data mining techniques in customer relationship management: A literature review and classification." *Expert systems with applications* 36(2):2592–2602.

- Niculescu-Mizil, Alexandru, Claudia Perlich, Grzegorz Swirszcz, Vikas Sindhwani, Yan Liu, Prem Melville, Dong Wang, Jing Xiao, Jianying Hu, Moninder Singh et al. 2009. "Winning the KDD cup orange challenge with ensemble selection." *The 2009 Knowledge Discovery in Data Competition (KDD Cup 2009) Challenges in Machine Learning, Volume 3* p. 21.
- Olle, Georges D Olle and Shuqin Cai. 2014. "A Hybrid Churn Prediction Model in Mobile Telecommunication Industry." *International Journal of e-Education, e-Business, e-Management and e-Learning* 4(1):55.
- Owczarczuk, Marcin. 2010. "Churn models for prepaid customers in the cellular telecommunication industry using large data marts." *Expert Systems with Applications* 37(6):4710–4712.
- Parvatiyar, A. Sheth, J. N. 2001. "Customer relationship management: Emerging practice, process, and discipline." *Journal of Economic Social Research* 23(3):1–34.
- Pendharkar, Parag C. 2009. "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services." *Expert Systems with Applications* 36(3):6714–6720.
- Prasad, U Devi and S Madhavi. 2012. "Prediction of churn behavior of bank customers using data mining tools." *Business Intelligence Journal* 5(1):96–101.
- Richter, Yossi, Elad Yom-Tov and Noam Slonim. 2010. Predicting Customer Churn in Mobile Networks through Analysis of Social Groups. In *SDM*. Vol. 2010 SIAM pp. 732–741.
- Rygielski, Chris, Jyun-Cheng Wang and David C Yen. 2002. "Data mining techniques for customer relationship management." *Technology in society* 24(4):483–502.
- Saradhi, V Vijaya and Girish Keshav Palshikar. 2011. "Employee churn prediction." *Expert Systems with Applications* 38(3):1999–2006.
- Sharma, Anuj, Dr Panigrahi and Prabin Kumar. 2013. "A neural network based approach for predicting customer churn in cellular network services." *arXiv preprint arXiv:1309.3945*.
- Swift, R. S. 2001. *Accelerating customer relationships: Using CRM and relationship technologies*. Prentice Hall PTR.

- Thearling, K. 1999. "An introduction of data mining." Direct Marketing Magazine.
- TODEREAN, Ionut B BRANDUSOIU Gavril. N.d. "PREDICTING CHURN IN MOBILE TELECOMMUNICATIONS INDUSTRY." . Forthcoming.
- Tsai, Chih-Fong and Mao-Yuan Chen. 2010. "Variable selection by association rules for customer churn prediction of multimedia on demand." *Expert Systems with Applications* 37(3):2006–2015.
- Tsai, Chih-Fong and Yu-Hsin Lu. 2009. "Customer churn prediction by hybrid neural networks." *Expert Systems with Applications* 36(10):12547–12553.
- Vafeiadis, Thanasis, Konstantinos I Diamantaras, G Sarigiannidis and K Ch Chatzisavvas. 2015. "A comparison of machine learning techniques for customer churn prediction." *Simulation Modelling Practice and Theory* 55:1–9.
- Van den Poel, Dirk and Bart Lariviere. 2004. "Customer attrition analysis for financial services using proportional hazard models." *European journal of operational research* 157(1):196–217.
- Verbeke, Wouter, David Martens, Christophe Mues and Bart Baesens. 2011. "Building comprehensible customer churn prediction models with advanced rule induction techniques." *Expert Systems with Applications* 38(3):2354–2364.
- Verbeke, Wouter, Karel Dejaeger, David Martens, Joon Hur and Bart Baesens. 2012. "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach." *European Journal of Operational Research* 218(1):211–229.
- Wang, Yi-Fan, Ding-An Chiang, Mei-Hua Hsu, Cheng-Jung Lin and I-Long Lin. 2009. "A recommender system to avoid customer churn: A case study." *Expert Systems with Applications* 36(4):8071–8075.
- Wei, Chih-Ping and I-Tang Chiu. 2002. "Turning telecommunications call details to churn prediction: a data mining approach." *Expert systems with applications* 23(2):103–112.
- Xia, Guo-en and Wei-dong Jin. 2008. "Model of customer churn prediction on support vector machine." *Systems Engineering-Theory & Practice* 28(1):71–77.
- Xie, Yaya, Xiu Li, EWT Ngai and Weiyun Ying. 2009. "Customer churn prediction using improved balanced random forests." *Expert Systems with Applications* 36(3):5445–5449.

- Yu, Xiaobing, Shunsheng Guo, Jun Guo and Xiaorong Huang. 2011. “An extended support vector machine forecasting framework for customer churn in e-commerce.” *Expert Systems with Applications* 38(3):1425–1430.
- Zhao, Yu, Bing Li, Xiu Li, Wenhuan Liu and Shouju Ren. 2005. Customer churn prediction using improved one-class support vector machine. In *Advanced data mining and applications*. Springer pp. 300–306.

Apêndice