Will You Stay or Will You Go?

Svante Pagels

DEPARTMENT OF DESIGN SCIENCES
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY
2019

MASTER THESIS

Will You Stay or Will You Go?

Churn Prediction for an App-Delivered International Calling Service

Svante Pagels

LUNDS UNIVERSITET

Will You Stay or Will You Go?

Churn Prediction for an App-Delivered International Calling Service

Published by

Department of Design Sciences
Faculty of Engineering LTH, Lund University
P.O. Box 118, SE-221 00 Lund, Sweden

Subject: Innovation Engineering (INTM01)
Division: Innovation Engineering
Supervisor: Lars Bengtsson, Division of Innovation Engineering, Department of Design Sciences, Faculty of Engineering LTH, Lund University
Examiner: Malin Olander, Division of Innovation Engineering, Department of Design Sciences, Faculty of Engineering LTH, Lund University

Abstract

In markets with fierce competition and low switching costs, the capability to keep customers is just as important as the capability to get new customers. One of the cornerstones of efficiently keeping customers from churning is the ability to identify customers that are more prone to leaving the company than others. This way, churn prevention resources and measures can be efficiently directed towards the customer segments where the impact is the greatest.

This paper investigates how churn can be predicted for an individual customer in an app-delivered international calling service (Rebtel). It also investigates which the underlying behavioral drivers of churn are and how these findings can be used to efficiently prolong the lifetime for customers using Rebtel's subscription services when calling to India. The method applied to identify churn is the machine learning algorithm Random Forest.

It is found that the Random Forest algorithm is well suited for classifying which customers that are about to churn by identifying complex customer behavior patterns indicating that a certain customer is about to leave the service. When investigating individual customer features to isolate drivers of churn, the method of using the Random Forest algorithm is not the best suited approach and contributes with limited findings.

The recommendations to Rebtel is that a churn prediction model implementing the Random Forest algorithm should be used to segment which users, among users calling to India, that should be incentivized to stay as customers in order to decrease churn.

Keywords: Retention, Churn Prediction, Customer Relationship Management, Machine Learning, Random Forest

Sammanfattning

På marknader med hård konkurrens och låga kostnader för att byta leverantör är förmågan att behålla kunder lika viktig som förmågan att skaffa sig nya kunder. En av hörnstenarna för att effektivt undvika att kunder lämnar, är att kunna identifiera vilka kunder som är mer benägna att lämna företaget än andra. På så sätt kan åtgärder för att motivera kunder till att stanna riktas mot kundsegment där effektiviteten är störst.

Detta arbete undersöker hur man kan prediktera vilka kunder som kommer att lämna en app-levererad internationell samtalstjänst (Rebtel). Kundattribut och beteenden undersöks även för att finna underliggande anledningar till att kunder lämnar tjänsten. Slutligen visas hur dessa resultat kan användas för att effektivt förlänga livslängden för kunder som använder Rebtels abonnemangstjänster för att ringa till Indien. Metoden som används för att identifiera vilka kunder som kommer att lämna tjänsten är maskininlärningsalgoritmen Random Forest.

Det konstateras att Random Forest-algoritmen är väl lämpad för att klassificera vilka kunder som sannolikt kommer att lämna genom att identifiera komplexa beteendemönster, specifika för kunder som säger upp tjänsten. Random Forest-algoritmen är dock inte särskilt lämpad för att undersöka underliggande drivkrafter hos kunder som lämnar och bidrar i detta område endast till begränsade insikter.

Rekommendationen till Rebtel är att använda sig av en klassificeringsmodell som implementerar Random Forest-algoritmen för att segmentera vilka av användarna som ringer till Indien, som bör approcheras med incitament eller andra åtgärder för att se till att de stannar.

Acknowledgments

I would like to direct a big thanks to my supervisor Lars Bengtsson, who has contributed with great advices in how to navigate the academic jungle as well as interesting discussions on the topic of loyal customers. Also, I would like to thank Jessica Wadin and Emil Åkesson for the additional coaching.

I would also like to thank my colleagues at Rebtel for always being curious when it comes to my thesis work and for taking the time to answer my many questions.

Stockholm, January 2020

Svante Pagels

Table of contents

List of Figures

List of Tables

1 Introduction

This section describes the background, problematization, aim and research question of this thesis. It also puts forward how this work will contribute to current research and the delimitations of the subject.

1.1 Background

Around year 2005 many tech companies faced a common problem. They had spent the last years pouring resources into customer acquisition and had a steady stream of new customers using their products, but the user base was not growing as much anymore. The reason for this was that while new customers were coming in, a substantial part of the users that were onboarded some while ago were leaving the services (Atkins, Gupta and Roche, 2018). A more recent example of the effects of poor retention is the home cleaning start-up Homejoy, a Sillicon Valley company that in short time managed to raise $64 million in initial funding. Through an aggressive promotional strategy, the company managed to acquire a large customer base early, only to have the whole enterprise demise when customers left the service shortly after (Robehmed, 2014; Farr, 2015). This phenomenon is not only prominent in the tech scene but occur in many markets with fierce competition and continuous introductions of new product offerings.

In the 1980s, researchers were trying to explain the importance of keeping customers and how this could be done. Some contributors were Berry, (1983) by introducing the term *relationship marketing*, a marketing strategy that emphasized customer retention and satisfaction, Gummesson, (1987) who emphasized the long-term implications of building and maintaining relationships with customers. Although these studies had little impact on the contemporary ways of doing business (Ballantyne, Christopher and Payne, 2003), they have laid the foundations of areas such as modern customer relationship management (CRM) and churn management.

In the late 2000s, more and more companies started to shift their focus from an offer-centric strategy with primary objective to bring in as many new users as possible, to a customer-oriented approach that focuses on keeping current customers by reducing the churn (Blattberg, Kim and Neslin, 2008). A study conducted by Reinartz, Thomas, & Kumar, (2005) looked into the balance between resources going toward customer acquisition and customer retention. They found that underspending on retention had a greater negative impact on long-term profitability than underspending on acquisition. The benefits that are derived from increasing customer lifetime by increasing retention (decreasing churn) are, among others, lower acquisition costs, revenue growth and lower operating costs (Reichheld, 1996).

However, keeping customers is not an easy task and often requires spending on retention campaigns, discounts or product development that prevents customers from becoming dissatisfied with the product. Therefore, it is crucial that CRM or discounts efforts are aimed to the right customers and product development focuses on the right feature improvements. Thus, a first step in any such retention effort is finding out who to target with retention campaigns (Buckinx and Van Den Poel, 2005) or what steps to take to improve the product by looking at the behavior that churned users exhibit.

Which brings us to the topic of churn modelling or churn prediction. The first mention of churn prediction in an academic setting is from an article published in 1991 where Empirical Bayes methods was applied to model customer churn in order to forecast circuit activity on a telephone network (Greis and Gilstein, 1991). Although the intention of that study was not to detect churn in order to increase revenue, others soon started to model churn in the increasingly competitive telecommunication (telco) industry for the purpose of helping companies retaining their customers (Madden, Savage and Coble-Neal, 1999; Wei and Chiu, 2002; Au, Chan and Yao, 2003; Shin and Sohn, 2004; Hung, Yen and Wang, 2006). During this period, the telco industry was experiencing a wave of liberalization in many parts of the world and the increased competition forced companies into innovative ways of keeping their customers. By using data collected on the customers such as demographic data, contractual data, customer service logs and call patterns, researchers employed classification techniques in order to predict which customers that were likely to switch to a competitor (Wei and Chiu, 2002).

With time, researchers have turned to ever more sophisticated classification techniques such as Decision Trees, Artificial Neural Networks, K-Nearest Neighbors, and Support Vector Machine (Keramati *et al.*, 2014) as well as modelling call patterns through Social Network analytics (Óskarsdóttir *et al.*, 2017). The application of churn prediction has also moved into new industries such as Mobile Games (Castro and Tsuzuki, 2015; Periáñez *et al.*, 2016; Milošević, Živić and Andjelković, 2017), e-commerce (Yu *et al.*, 2011), banking (Mutanen, 2006) and general subscription and non-subscription settings (Jahromi *et al.*, 2010).

The ability to identify churners in a customer base, relies on these customers exhibiting a distinguishing behavior and that this behavior can be gathered from the data collected by the company on the customer. In telco settings, the features and models used to identify churn have been evaluated in a long row of academic papers (Madden, Savage and Coble-Neal, 1999; Wei and Chiu, 2002; Shin and Sohn, 2004; Hung, Yen and Wang, 2006; Jahromi *et al.*, 2010; Keramati *et al.*, 2014; Hughes, 2015; Óskarsdóttir *et al.*, 2017; De Caigny, Coussement and W. De Bock, 2018; Mitrović *et al.*, 2018). Also in Software-as-a-Service (SaaS) and e-commerce settings, although not as prevailing, some scholars have looked into the features and models used to identify churners (Frank and Pittges, 2009; Periáñez *et al.*, 2016; Subramanya and Somani, 2017).

Although, different features and models have been presented in previous research there is no generic model for churn prediction and previous applications of models and feature selections are industry or even company specific.

Other than pointing out which customers that are the most likely to churn, some classifying algorithms (classifiers) also have the added benefit of giving the importance of different features when determining the probability of a customer churning. An examples of classifiers providing this importance score are Decision Tree methods (Kohavi and John, 1997). Such *feature importance* can be used to identify the factors that are most important to keeping a customer and how to continue developing the product or service. (Ivanonv, 2018)

1.2 Problematization

Rebtel Networks AB (Rebtel) is a Swedish company specializing on communication and financial services for immigrant groups around the world. Their main products can be divided into two categories; calling and financial products. Within the calling products the customers can either buy “World Credits” which can be used to call destinations all over the world or buy a plan which provides unlimited or limited calling to certain destinations for a monthly fee. The financial products are primarily Mobile Recharge products which allows the customer to refill the prepaid credit on foreign prepaid telephone plans, primarily in Cuba. The company delivers its services through their app, which is available on iOS and Android devices, and through a web portal. Customers are scattered all over the world and some of the most prominent ethnic groups that use Rebtel to call to their homeland are Indian, Cuban, Nigerian and Bangladesh. Even though Rebtel solves a similar need for each of these customer groups – the need to communicate with friends and family in the home country, they all exhibit different behaviors in usage and loyalty.

In the last couple of years, Rebtel has put a lot of focus into developing their offering and attracting new customers. This focus has led to a substantial growth in customer acquisition and revenue (UC Allabolag AB, 2018). Recently though, the company has changed its focus from an offer-centric to a customer-oriented approach, where one of the main strategic goals for 2018 is to increase the customer lifetime by decreasing churn. As described in the previous section, one of the cornerstones of effective churn management operations is the ability to efficiently identify the customers that are more prone to churning than others (Buckinx and Van Den Poel, 2005).

In the case of Rebtel, customers exhibit very different behaviors depending on where in customer journey they are. Just as for other app providers, Rebtel has a high churn rate for new customers (Statista, 2018b), while later in the customer journey the churn is substantially lower. To efficiently identify drivers of churn in this setting it is important to specify which segment that is being analyzed. In the initial phase of the customer journey, mainly the period from signing up as a customer until a couple of weeks in, the reasons for churn have been thoroughly analyzed at Rebtel. Although, few efforts have been done to try and identify customers that are more likely to churn later in the customer journey.

Even though a lot of studies have been made regarding the drivers of churn and churn prediction, few or none address how to approach this in a communications company with an app as their product. The fast-paced, global and highly competitive market with low switching costs that is the app market is much more affected of churn than the, mostly, subscription-based mobile communications market (Statista, 2018a, 2018b) and the retention mechanisms are a lot less investigated.

1.2.1 The International Calling Market

Rebtel's main market, international calling, is an industry that has been through several systemic changes in the last few decades. In the beginning of the 1990s, an U.S. international call cost on average $1/minute and 67 % of the international long distance market revenue came from businesses. (Telegeography, 2017) The coming decade brought a wave of market liberalization all over the globe which in turn brought new market entrants and many people saw the price of both domestic and international calling plummeting due to increased competition. In the beginning of the 2000s, the price of cheap cellphones had gone down and the cellphone penetration in countries with high emigration increased rapidly. All of this caused a shift in the source of international calling revenues from business calls to social calling and by 2015 calls to mobile phones in developing countries accounted for 65 % of all international calling traffic. This growth of social calling fueled a boom in the international calling market that lasted for almost 20 years. Although it has declined in recent years, the international calling market is estimated to generate $70 billion in yearly revenues. (Telegeography, 2017)

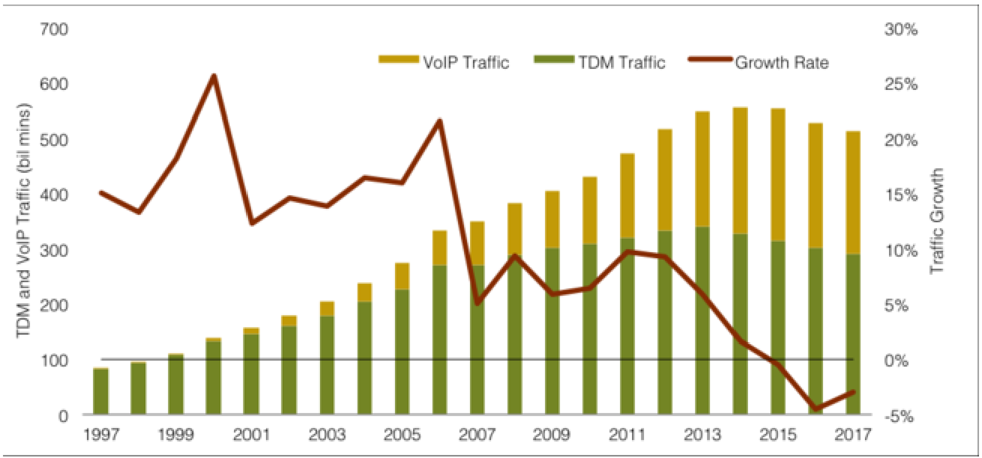

Figure 1: International calling market, excluding pure VoIP calls, such as Apple Factime, Google Duo, Facebook Messenger and Whatsapp. (Telegeography, 2017)

This growth and price competition left the traditional telco actors vulnerable to new technology in a world where high-speed internet penetration increase with 24 % annually in developing countries. (ICT Data and Statistics Division, 2017)

Starting with Skype's launch for desktop devices in 2003, VoIP applications for smartphones such as WhatsApp, Facebook Messenger, WeChat, QQ, Viber, Line, and KakaoTalk have an estimated combined user base of over 5 billion in September 2017 and pose a real threat for the traditional companies in the market. (Telegeography, 2017) The entrance of these actors can be seen in the decline in growth of the traditional international calling market in *Figure 1*.

In these turbulent market conditions, Rebtel has found a niche of providing calling with high quality to countries with unreliable or unavailable internet as an alternative to having to rely on a defective internet connection. Even though this tactic has worked well, Rebtel is not immune towards the general market trend and are aware that they will have to work hard to keep their customers from switching entirely to a cheaper internet-to-internet calling provider such as the ones mentioned above.

1.3 Aim and Research Questions

The aim of this study is to build a model that identifies customers that have a high probability of churning while also determining some key drivers of churn. This will be done for Rebtel's largest customer group, customers using Rebtel's subscription services for calling to India. More specifically, the research questions are:

1. How can individual customer churn be predicted in an app-delivered international calling service?
2. What are the drivers of churn during the medium and long-term retention period for an app-delivered international calling service?
3. What should Rebtel do to improve the medium phase and long-term retention?

1.4 Delimitations

To make this thesis practically viable, some delimitations are inevitable. The delimitations used in this paper are the following:

- The features used to model churn come exclusively from Rebtel's internal data on the customers, no data collected from other sources will be used in the model. The reason for this is to ensure that the final model is easily replicable and can be used for future business applications.
- In a business setting it is crucial to estimate the value of a customer as well as the probability of them churning before engaging in measures to retain them. This is to ensure that resources are not directed at customers that will not provide the firm with any additional value. This thesis will not look at the customer value, instead only focus on the probability of individual customer churn and the underlying reasons behind it.
- Retention behavior can be observed during different periods of the customer relationship life cycle; initial retention period, medium retention period and long-term retention. The churn that will be modeled in this paper is churn occurring in the medium retention period and long-term retention. An elaboration regarding the differences of these definitions can be found in section *2.1 Customer Retention and Customer Relationship Lifecycle.*
- Only subscription users with India as their main destination will be examined in this paper. The reason for not including other products or calling destinations is that behavior is believed to be significantly different between different destinations and between pay-as-you-go products and a single model will only be able to capture very general relationships in such a setting. The reason for choosing subscription products to India as the area to further investigate is that is the most substantial product-destination combination at Rebtel with regards to revenue, margin and user base.

1.5 Thesis Outline

Chapter 1: Introduction
This chapter introduces the reader to the background of the thesis, the industry and company being investigated and gives a clear description of what the objective of the thesis is.

Chapter 2: Theoretical Framework
The theoretical framework explains previous research within the areas of customer retention, churn prediction and the algorithms employed to identify potential churners through machine learning. This is the result of the performed literature review.

Chapter 3: Method
In this chapter the methodology and research characteristics of the paper is explained.

Chapter 4: Empirical Findings
Based on the literature review, data collection and modelling described in the previous chapter, this chapter presents the empirical findings from the final model.

Chapter 5: Analysis and Discussion
The analysis and discussion chapter goes through the empirical findings and reflects on the findings with regards to the research questions. The metrics of the classification algorithm are also investigated to validate the model, in order to answer *Research Question 1*.

Chapter 6: Recommendations and Conclusion
Finally, the findings are concentrated into recommendations to Rebtel and overall conclusions regarding *Research Question 1* and *Research Question 2*.

2 Theoretical Framework

This section describes previous research within the relevant areas of this thesis. It aims to outline the concepts of why customer retention is important, customer churn and churn modelling.

2.1 Customer Retention and Customer Relationship Lifecycle

Companies that move from an offer-centric business model to a customer-oriented one, do so to change the relationship with their customers. By focusing on fulfilling the customers need the company is hoping to add more value than the competitors and not only be judged on the price of the products or services. This shift in customer mentality from a *transaction-orientation* to a *relationship-orientation* has been described by Christian Grönroos (2008). According to Grönroos this shift has several benefits from the customer's perspective. By establishing a relationship with the company, the customer gets an increased sense of security and trust to the company, increased recognition and opportunity to get beneficial added services. All these factors increase the customer's propensity to stay as a customer.

Grönroos also propose a customer relationship life cycle where every customer relationship to a company can be divided into three phases. During the *Initial stage* the customer does not know about the company's service. If the customer subsequently gets to know about a service that the company provides and that might satisfy a customer need, the customer enters the *Purchasing process*. The purchasing process is when the customer evaluates the service and what it is worth. Finally, the customer enters the *Consumption process*, when he or she makes the buying decision and gets to use and evaluate the service and to which degree it satisfies the customer needs. In-between these steps the customer must decide whether to move on to the next step and stay in the relationship life cycle or to end the relationship. When the consumption process is finished the customer can choose to restart the journey by repurchasing a service from the company. In each of these phases specific activities and targets should be pursued from the company side in order to build a positive relationship with its customers. (Christian Grönroos, 2008)

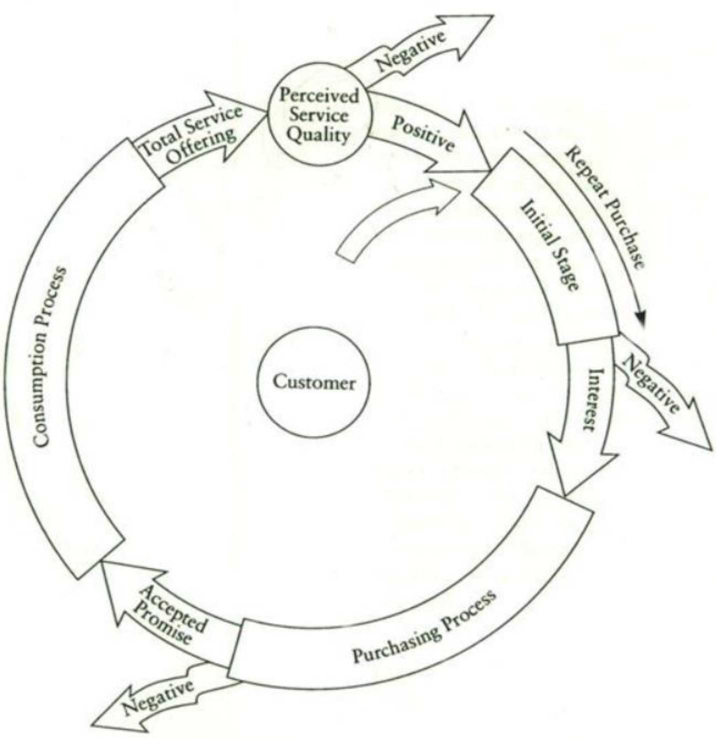

Figure 2: Customer relationship lifecycle (Christian Grönroos, 2008).

At Rebtel and similar tech companies providing a consumer service, a common practice during the purchasing process is to provide a free or heavily subsidized trial of the product to facilitate and nudge the customer towards a positive purchasing decision. At the point when the customer must go from a free or subsidized version of the service to paying the full price, many opt out. In the tech scene, this early phase of the customer-company relationship has been described as *initial retention period* in retention contexts. The period during the consumption phase is referred to *medium retention period* and the period where new features or services must be provided to the customer in order to get them to re-purchase is referred to as *long-term retention*. (Ellis and Brown, 2017)

In the case of Rebtel, the initial retention period can be seen as the first month after the user starts a trial, at Rebtel referred to as *Welcome Offer*. During this first month the churn figures are higher than for the rest of the customer journey and as Ellis and Brown explain in their book from 2017 the retention during this initial period has mostly to do with the stickiness of the product. For the medium retention period the reasons for churning are different. Here a customer will be lost if the service or product fails in becoming a natural part of the customers life, in other words if the customer does not build any habits around the provided service. The long term retention relies on the company’s ability to innovate and enhance the current features or add new features. (Ellis and Brown, 2017)

As described in the delimitations segment, this paper will focus on the medium and long-term retention periods since it is for these periods usage data will be available for each customer.

2.1.1 The Effect of Increased Retention

Often, the business value of retaining a customer is not obvious. The ordinary metrics to determine the successfulness of a business are often focused on current period costs and revenues and do not effectively give an indication of cash-flows to come. The customers’ lifetime on the other hand has an effect on the long-term cash-flow of the business. (Reichheld and Sasser, 1990) To best illustrate this issue imagine two different high growth firms:

Firm A has a yearly intake of 1000 new customers each year and a monthly churn of 7 %

Firm B has a yearly intake of 1000 new customers each year and a monthly churn of 4 %

What long-term effects will the higher churn of *Firm A* have on the company’s revenue?

Assuming they both start with a customer-base of 1000 customers and that each customer contributes with the same revenue, the increase in revenue will be the following for each firm.

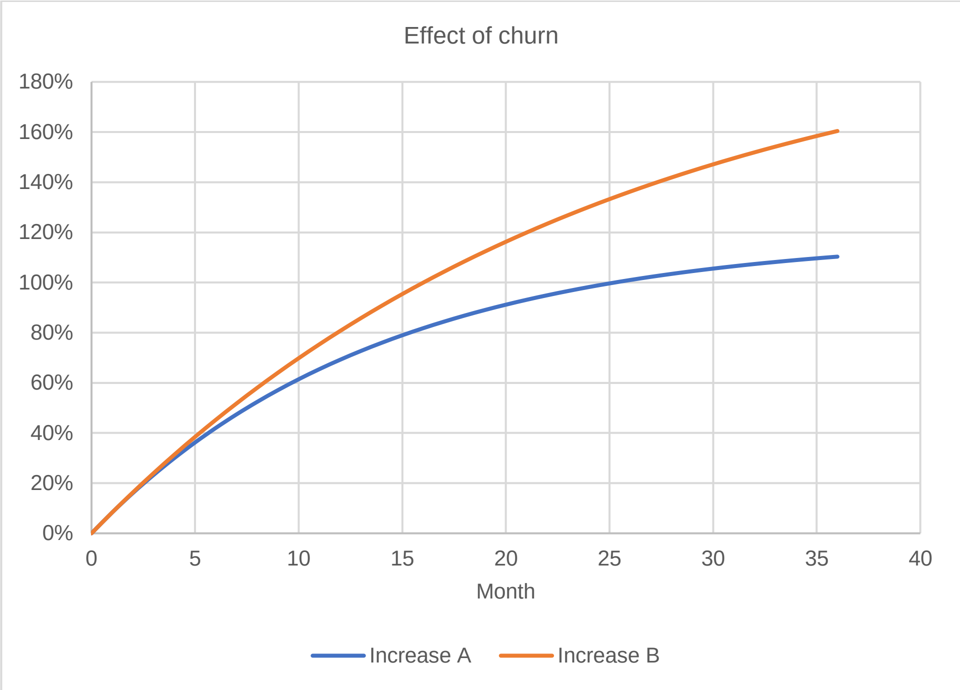

Figure 3: Effects of churn on two fictional firms.

By acquiring the equivalent of the whole customer base at month 0 yearly, *Firm A* manages to increase their revenues with 100 % after 25 months. *Firm B* on the other hand, which has a much lower churn rate manages to do this in only 16 months. After three years, *Firm A* has increased its revenues with about 110 % while *Firm B* has increased the revenues with 160 %. As can be seen in the graph above, *Firm A* is barely increasing its revenue after three years even if they are bringing in new customers. This is referred to as "flatlining" by the tech company Hubspot's previous VP Growth (Balfour, 2015). As this example illustrates, churn can greatly impact a business and its growth.

In an article in the Harvard Business Review, Reichheld and Sasser refers to the loss of potential revenue growth as the cost of losing a customer (Reichheld and Sasser, 1990). Not only will a lower churn result in a higher revenue just through the increase of the customer base, the cost of losing a customer is also impacted by several other positive effects of retaining long-term customers.

As customers become accustomed to using a certain product or service and are satisfied, they increase their spending. Reichheld and Sasser found that this was true for each than more of 100 different companies operating in 24 different industries. They also found that as the customers purchases rose, the operating costs related to these customers declined. According to the authors this has mainly to do with two factors; as the tenure of the customers increases the customers learn how to use the products or service, have less question and the company better understands how to serve said customers cost efficiently. Also, companies with a long relationship with their customers have been found to be able to charge more for their products or services as well as getting free advertising through the word-of-mouth effect of long-lived customers. (Reichheld and Sasser, 1990)

The financial effects of these phenomena are staggering. Reichheld and Sasser estimate that a 5 % rise in customer retention increase profits by 25-85 %. In the mobile communication sector it has been estimated that a 30 % decrease in churn will result in a 15 % increase in long term revenues (Bolton, 1998). Also for shareholders the retention rate is something that should be important. According to Gupta, Lehmann and Stuart, (2004) a 1 % improvement in retention increases firm value by 5 %.

Today, a lot of the new knowledge around customer retention and churn prevention is coming out of Silicon Valley and the tech industry. Companies such as Evernote and Hubspot have gained market leading positions to a great part by having a stated focus on decreasing churn and developing products that retain as many customers as possible. In the fast paced, low entry-barrier market that many tech firms find themselves in, a business strategy that builds on keeping the customers that are acquired has proven very successful. (Ellis and Brown, 2017)

2.2 Churn

Since the *churn* expression is central to this paper, this section aims to explain it in further detail. The main purpose of the following sections is to specify what churn is and how it is defined in this paper.

2.2.1 Different Types of Churn

Of course, there are many different reasons that a customer can have for leaving a service or product. But when talking about churn one general categorization can be done; involuntary churn or voluntary churn. Voluntary churn is defined as when the end user actively decides to quit using the service, often by clicking an unsubscribe button or alike. Involuntary churn is when a customer does not actively choose to leave a service. Reasons for involuntary churn are often things like a subscription ending, relocation or payment issues. (Cheney, 2016)

In the case of Rebtel, the distinction between involuntary churn and voluntary churn is not very useful. Due to the nature of the pay-as-you-go product they offer, customers can churn simply by not making any calls which would be defined as voluntary churn. Whereas for their limited or unlimited products which are subscription-based and requires the customer to actively unsubscribe or remove payment details to stop monthly payments which is a clear case of voluntary churn. There are also cases of involuntary churn, primarily due to discontinuation of products, although no such products are included in this paper.

2.2.2 Defining Churn

Churn, in the general sense is defined as the loss of customer ('Churn', 2018). At first glance, it seems like a measurable and well-defined metric but as one digs deeper it is evident that a more stringent definition is needed. Previous studies that have looked into churn in an non-contractual setting where no active measure is needed to stop being a user, such as the pay-as-you-go product offered by Rebtel, all face this problem and in order to build a predictive model the first step is to develop a sensible definition of churn (Jahromi *et al.*, 2010).

The most intuitive way of defining churn is when a customer has not been active for a certain period of time. Although this poses a new question, which period should be used as cutoff? If a very short period of time is chosen, say 7 days, many non-frequent users would be falsely flagged as churned users. If a long period of time is chosen, say 90 days, the resulting predictive model may be too bland on customers, missing churners that have churned and subsequently returned to the service. Other researchers have attempted to solve this by dividing the customer base into different segments and applying different cutoff for each segment (Jahromi *et al.*, 2010). Although this methodology has worked in other settings it has the downside of increasing the complexity of the churn definition and making the results less interpretable. If the churn concept means different things for each customer segment, one must treat it segment differently if using the results in a business setting.

The customer base at Rebtel consists of both contractual customers, who have purchased a limited or unlimited calling plan, pay-as-you-go users who buy credits which can be used for calling abroad, money transfer customers who use Rebtel services to send money abroad and combinations of the above. Therefore, a definition of churn must be defined in a way that work in all of the different customer settings.

In this paper, Rebtel's internal definition of churn will be used:

'A customer is churned when no **payment** or **action** is performed in 30 days.'

In the case of contractual customers, it is not necessary that they make calls, or use any other services, to be considered active customers, since they pay a monthly fee for the services of Rebtel. As long as a contractual customer has an active calling plan with Rebtel the customer will be considered active. In the case of pay-as-you-go customers only looking at payments does not work since you can fill up your account with a large amount of calling credits and then use them for more than 30 days. Therefore, the definition also states that any action within 30 days will deem a customer active. The term *action* will in this paper refer to a call being made.

A monthly payment for a subscription service is internally viewed as 30 small daily payments, as revenue for the payment is recognized during the full period of the products' validity. This means that if a payment is performed on day 1 and the customer performs no other action, the customer will be considered churned on day 60 – 30 days after the last revenue recognition of the payment.

Since this paper only investigates churn for subscription customers calling to India, users that continues to call with India as destination but does so with a pay-as-you-go product will also be considered as churned. The reason for this is that they have left the service that is being investigated and should therefore be included when investigating why customers leave.

2.2.3 Churn and Inactivity

Although the above definition of churn generally does a good job of separating active users from churned ones it has one downside. Due to the behavior of many of the customers at Rebtel who only use the service sporadically they might be wrongly identified as churned if the frequency between their actions is more than thirty days. Such a user would first be identified as churned and later identified as a reactivated user using Rebtel's internal definitions, even though the customer never had the intention of quitting the service. An example of such users are the customers who only use Rebtel to send mobile recharge to Cuba. They send these recharges on an average once every 60 days. Also, Rebtel operates in a highly competitive industry with low switching costs and constant price competition. Therefore, many users repeatedly switch between or use several providers. This can also wrongfully label a customer as churned, when he or she in fact may still have the intentions of continuing using the service.

In academia this topic has been discussed previously. In an issue of the Harvard Business Review, author Barbara Bund Jackson makes the distinction between *lost-for-good* customers and *always-a-share* customers. The lost-for-good customers often appear in industries with high switching costs and with customers that are reluctant to switch between different providers. Whereas the always-a-share customers are prevailing in industries with low switching costs where customers change providers often or even might be using several services at the same time. (Jackson, 1985)

Although this definition can be used to separate actual churned users from infrequent users due to usage of competitors services, it doesn't help with separating churned users from users that simply use services such as the ones supplied by Rebtel occasionally.

To address this issue we turn to the academics who have investigated customer lifetime value, where the question "Is a certain customer active or not?" is crucial. Reinartz and Kumar, (2000) showed that the probability of a customer being active can be derived from three measurements; total time since beginning of being a customer, time since last purchase and number of purchases. The findings can be summarized as an active customer is someone who displays frequent and regular interactions with a company.

Since the reason for being inactive is not important, but the fact if a user has churned or is simply displaying their ordinary low frequency behavior is, this paper will distinguish between churned and low activity, or *dormant*, users simply by their previous behavior. A distinction between a churned user and a dormant user which draws from both the work of Jackson and Reinartz and Kumar, was formalized by Crié, (2002). It states that a churned user is someone that has been a customer but currently has no likelihood of undertaking new purchases in the near future. A dormant customer on the other hand is a customer who, taking into account his or her usual buying pattern shows a weak probability of repeating a purchase in the near future.

By using this definition, dormant users can be excluded from the modeling which means that the model will become better at identifying actual churned users.

2.3 Churn Prediction

Hung, Yen and Wang (2006) describes that 'churn management' first appeared in academia in the beginning of the 2000's. The practice of proactively working to keep customers by using data modelling techniques to predict their propensity to leave the service, was concentrated to the telco industry and grabbed the attention of many scholars. 'Churn management' relies on models that accurately predict a customer's decision to move from one operator to another, in this paper this modelling is referred to as c*hurn prediction*. This section describes some of the most common techniques and best practices on how this is achieved.

2.3.1 Data Mining

Data mining is defined by the SAS Institute as a process to look for hidden patterns in data that can be used to predict future behavior. The process consists of several steps and it starts with asking a business question that you want to answer. Based on that question the analysis progresses with collecting and exploring available data. Although data is more available and accessible than ever before, the process of collecting and exploring this data is often very time consuming. Data often need to be joined from different sources and raw data might have to be transformed into a form that can be used as input for the coming steps of the data mining. There is also a need to choose among data what to include or not. Choosing which data to include in the model often demands knowledge of the scenario that is being modelled and a great deal of visualization of relationships and testing. (SAS Institute, 2016)

One test that is commonly used when choosing which data to include in prediction modelling is to test continuous independent variables against the dependent variable with a Z-test. The Z-test compares the feature distribution among customers who churn and customers who are retained and if there is a significant difference between the two groups, the feature is kept. (Hung, Yen and Wang, 2006)

When it comes to categorical values such as *Has made a call in the last 3 days* the approach of using Z-test is not viable since there is no continuous value to look at the distribution of. In this case a Chi-square test can be used to determine whether the churn variable is significantly dependent on a certain value. These statistical tests might be good at determining the usefulness of a certain feature but should not be used to explain to which degree the feature impacts the dependent variable. (Manning and Raghavan, 2009)

The exploration of the data is a way of gaining deeper understanding of which information that is stored in the data. By looking at correlations, trends and patterns, insights into how the data might be used to answer the *business question* from an analytical perspective are formed. This step might include some feature engineering where new variables are created by transforming data or combining different data.

When the exploration of the data is done the next step is the actual modelling. The goal of this step is to find a combination of data and prediction model that accurately represents the relationships in the data that answer the business question. This process heavily relies on experimentation and trial and error, where different combinations of data and model are tested and compared. From the winning model, the importance of certain features can be derived and the results are interpreted and acted on in the business context. (SAS Institute, 2016)

2.3.2 Prediction Techniques

When dealing with predictions in a business settings the three most used types of modelling are regression techniques, neural networks and decision tree techniques. (Hung, Yen and Wang, 2006) All of these different categories of prediction techniques have different strengths and drawbacks.

In a comparison of the most popular techniques from 2015, the authors compare different techniques from all of the above categories and evaluate them using precision, accuracy, recall, AUC and F1 scores. More about these evaluation methods can be read in the section bellow; *Evaluating Model Performance*. The authors find that decision tree methods and artificial neural networks perform the best when it comes to customer churn prediction. (Vafeiadis *et al.*, 2015)

Although most previous literature put focus on the predictive power of these methods, another factor to take into consideration is the interpretability of the results. (Verbeke *et al.*, 2011) Methods such as artificial neural networks have high predictive power but are black-box methods and do not supply insight into what the drivers of churn are. This limits the possibilities of further analysis and business actionability of the results. Decision tree methods on the other hand work in such a way that a feature importance can be derived by assessing which features that have the most predictive power when predicting if a customer churn or not.

The combination of predictive power of the model and interpretability of the results make decision tree techniques the preferable way to predict churn in this paper.

2.3.2.1 Random Forests

Random forests are an ensemble method that builds on decision trees and is often used for classification problems. During the training, it constructs a large number of individual decision trees that are trained on different subsets of the input data. The output class of the prediction is determined by taking the mode of the class predictions of each of the individual decision trees within the model (Ho, 1995) or, as in the python implementation of random forest used in this paper, an average of the probability outputs from each individual tree. (Scikit-learn, 2018) It has been shown that taking the average instead of the mode produce consistent classification as long as the base classifiers are consistent. (Biau, Devroye and Lugosi, 2008)

A decision tree classifier is a model that uses rules to divide a dataset into subsets in which the probability of belonging to one class or another is greater. Classification and Regression Tree (CART) analysis was first introduced in 1984 (Breiman, 1984) and while decision trees for regression and classification purposes have some similarities they also differ in certain ways. A classification decision tree starts with the variable and rule that best splits the data into two different subsets, each having as high probability as possible belonging to a specific class. (Rokach and Maimon, 2005) This is illustrated by a simple decision tree, classifying a dataset of persons into female and male based on height and weight in figure 4 below.

To choose by which feature the dataset should be split and which threshold that should be used, the CART algorithm uses *Gini impurity* as split criterion. (Breiman, 1984)

For a binary classifier the Gini impurity is calculated as $I = 1 - \sum_{i=1}^{J} p_i^2$ where $J \in \{0,1\}$ are the binary classes and p_i denotes the fraction of items labeled with class *i* in the set. A perfect split results in a Gini impurity of 0 and the worst Gini impurity possible is 0.5 when the 50 % of the cases belong to one class and 50 % belong to the other. The CART algorithm chooses the feature and split that will decrease the Gini impurity the most.

Height	Weight	Gender
175	82	male
163	72	female
173	75	female
182	82	male
185	90	male
175	85	female

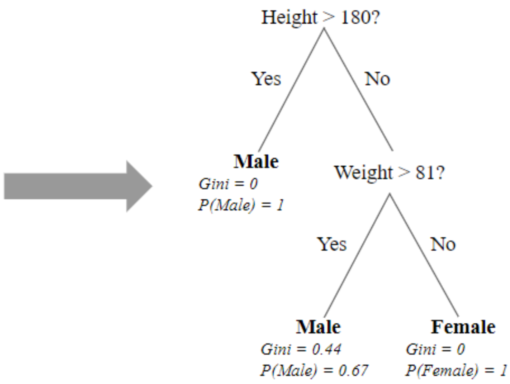

Figure 4: Example of decision tree.

Decision trees work great when classifying data they have been trained on, such as using the tree in figure 4 to predict the gender of the data in the table. The feature importance can also be derived by calculating the decrease in node impurity weighted by the probability of reaching that node and used to draw conclusions of which the most important features are. But, they have one big drawback, namely inaccuracy when it comes to new data (Hastie, Tibshirani and Friedman, 2009).

To use the interpretability from decision trees but with a vastly improved accuracy and robustness, the Random Forest classifier builds up a vast collection of tree classifiers, trained on different subsets of the data, and by taking the mode or average of each decision tree output, the output becomes more general and robust to noise in the dataset.

Formalized by Leo Brieman in a paper from 2001, a random forest can be described as a classifier consisting of a collection of tree-structured classifiers $\{h(\boldsymbol{x}, \Theta_k), k = 1, \ldots\}$. Where $h(\boldsymbol{x}, \Theta_k)$ represents the k:th decision tree, $\boldsymbol{x}$ is the input vector of independent variables and Θ_k is a random vector representing which parts of the input vector to be used to train the k:th decision tree. In the CART implementation of random forest used in the paper, Θ consists of a number of independent random integers between 1 and K. (Breiman, 2001)

The random forest algorithm does also have the great advantage of being very robust when noise or irrelevant variables are included in the dataset. (Hastie, Tibshirani and Friedman, 2009) This characteristic also makes it a good tool for feature selection, by adding all possible features in an initial model and reducing the number of features by only keeping features that show a high degree of importance.

2.3.3 Overcoming Imbalance

When using churn prediction to divide customers into two classes; churned or not churned, a common problem is that the dependent variable is imbalanced. (Burez and Van den Poel, 2009) This means that the fraction of customers who churn in a specific time period is often a lot smaller than the fraction of customers who do not churn. This presents a problem when building a model that is trying to classify each user into “churn” and “no churn” since a higher probability of a customer not churning makes the model more prone into classifying customers as “no churn”.

A simple example is when a model is presented with a training set of 100 users of which 5 churn. If the model is trained to classify the customers into a group of churners and a group of non-churners, it can simply put all customers in the group “no churn” and that way achieve the impressive accuracy of 95 %. Although, that model does not do a lot of good in a business setting. This phenomena is known as the *Accuracy Fallacy* and is further explained in the section *Recall, Precision, F1 Score and The Accuracy Fallacy*.

The problem of class imbalanced in churn datasets was investigated in a paper by Burez and Van den Poel (2009). They found that under-sampling the dataset lead to improved evaluation metrics, even though some information is lost due to not all of the data being used. The under-sampling process is a type of stratified sampling, where samples from the original dataset are chosen in such a way that the training dataset consists of as many churners as non-churners. Figure 5 illustrates this concept.

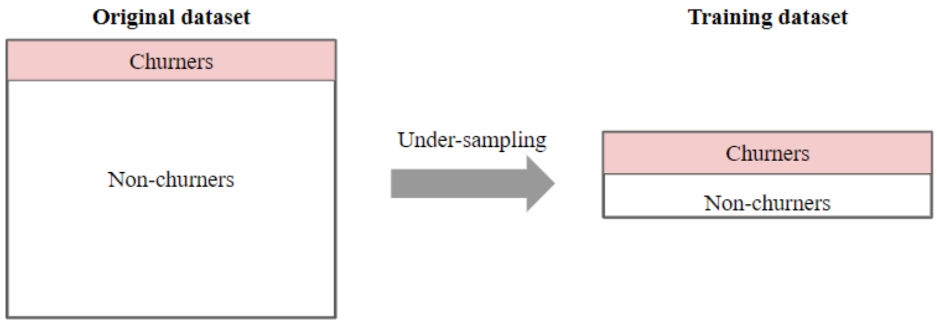

Figure 5: Under-sampling visualized.

2.3.4 K-Fold Stratified Cross-Validation

K-fold cross-validation is a procedure used to estimate the performance of the model on new, unseen data. It is a particularly efficient method to use on small datasets, although it is the preference for many people regardless of sample size. (Bramer, 2013)

The general procedure involves randomly dividing the dataset into k groups (folds), of approximately equal size. The first fold is used as validation set and the rest of the folds are used to train the model. This process is repeated k times, until all folds have been validation sets and all combinations of $k-1$ folds have been used for training. (James *et al.*, 2013)

If the dataset is imbalanced, the cross validation can be combined with a stratified sampling such as the one presented in the *2.3.3 Overcoming Imbalance* section above. As explained by Kohavi (1995), stratification within cross-validation is generally better that non-stratification since it leads to both lower bias and variance. It is important in this case that the validation fold remains imbalanced since it would otherwise result in overoptimistic results with regards to the model performance. (Blagus and Lusa, 2015) A schematic picture of the under-sampled cross-validation process is described in *Figure 6* bellow.

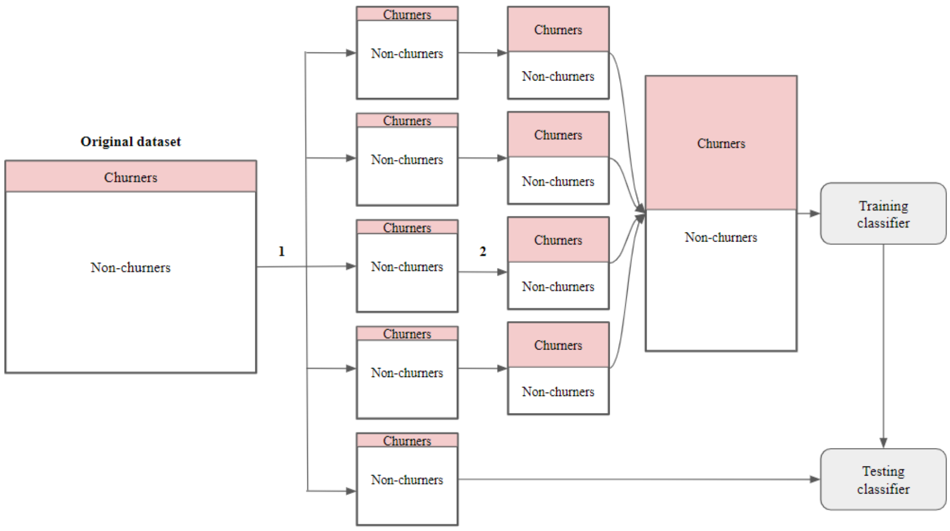

Figure 6: Stratified k-fold cross-validation

The dataset is first (1) split up into k folds. In the figure above $k = 5$. Each training set is then (2) under-sampled so that the dataset that is used to train the classifier has roughly the same number of churners as non-churners. The validation set is not under-sampled in order to get realistic test results of the classifier. This process is repeated k times so that all the folds are used as validation set exactly once.

2.3.5 Evaluating Model Performance

In order to be able to properly evaluate the performance of the classifier, several commonly used evaluation measures are introduced.

2.3.5.1 True Positives and False Positives

When working with binary classification problems such as churn prediction, true and false positives and negatives are the fundamentals of evaluating the classifier's performance. Firstly, a definition of what it means for an observation to belong to a positive or negative class. In our case, when trying to predict churn, we say that all observations that ended up churning belongs to the positive class. Correspondingly, all observed customers that did not churn belong to the negative class.

A true positive (TP) is an observation that the classifier rightly classified to belong in the positive class, while a false positive (FP) is an observation that the classifier wrongly classified to belong in the positive class. The same reasoning goes for negatives where you have true negatives (TN) and false negatives (FN). (Google Developers, 2018)

The ratio that the classifier successfully manages to classify the positive class is referred to as the true positive rate (TPR). While the ratio that the classifier does not manage to classify the positive class is referred to as the false positive rate (FPR).

2.3.5.2 Recall, Precision, F1 score and the Accuracy Paradox

While the true and false positives and negatives are good measurements, they do not contain much information by themselves individually. You can for instance not say that a model with more true positives is by definition better than a model with fewer true positives. An easy example of a model successfully identifies all members in the positive class is a model that predicts each observation to belong in the positive class. Such a model would not miss a single observation in the positive class but would most likely have a high number of false positives.

Therefore, in practice other measurements are commonly used to describe the performance of a classifier. Recall (r) is one of the measurements used, which is the ratio of the correct classified observations in the positive class to all the members in the positive class:

$$\mathbf{r} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{1. 1}$$

In effect this is the same as TPR and in applications where identification of positives is crucial, such as medical cases (you do not want to miss any patients with cancer), the recall is of great importance. (Powers, 2011)

Another commonly used measurement is the precision (p) measurement. Precision measures the proportions of predicted positives that are correctly classified:

$$\mathbf{p} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \tag{1. 2}$$

This is often used in machine learning, data mining and information retrieval and gives an indication of the accuracy of the predicted positives. (Powers, 2011)

In most applications though, just measuring the recall or the precision is not enough since a good model often requires a balance between the two. The F1 score is the harmonic average of the precision and recall:

$$\boldsymbol{F1} = \mathbf{2} * \frac{\boldsymbol{p*r}}{\boldsymbol{p+r}} \tag{1. 3}$$

The most widely used evaluation metrics though, is accuracy. Accuracy (a) measures the ratio of correct predictions to the total number of observations.

$$\boldsymbol{a} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \tag{1. 4}$$

In highly imbalanced datasets a model that has high accuracy can be worse than a model with lower accuracy, this is what is referred to as the *accuracy fallacy*. Despite optimizing classification error rate, models may fail to capture crucial information when classifying. (Valverde-Albacete and Peláez-Moreno, 2014) Since this report is dealing with a dataset that is imbalanced in its nature, the accuracy measurement will not be used primarily.

2.3.5.3 Receiving Operating Characteristic Curve

Receiving Operating Characteristics (ROC) analysis, is a measurement that originated in signal detection theory, but is now commonly used in machine learning applications. It has several advantages when comparing to the above-mentioned ways of analyzing classifiers performance. First of all, the ROC curve is a useful way of visualizing correct and incorrect classification across the entire range of class distribution. In other words, it works just as well for imbalanced data as balanced data.

The method goes about by plotting the TPR against the FPR, which are defined as:

$$\boldsymbol{TPR} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{1. 5}$$

$$\boldsymbol{FPR} = \frac{\boldsymbol{FP}}{\boldsymbol{FP+TN}} \tag{1. 6}$$

A classifier that only guesses randomly which class a certain observation belongs to would have equal FPR and TPR and would therefore produce a diagonal line from $(0,0)$ to $(1,1)$. A classifier that performs better than random guesses would have a higher TPR than FPR, but different for each decision threshold, T, that is used to classify a certain observation as belonging to the positive class or not. Consider a classifier with two classes “+” and “-“, that classifies a certain observation x_i to belong to a certain class depending on the decision threshold T:

$$\forall \boldsymbol{i} \quad \begin{cases} \boldsymbol{x_i} \to \{+\} \text{ if } \boldsymbol{P}(+|\boldsymbol{x_i}) \geq \boldsymbol{T} \\ \boldsymbol{x_i} \to \{-\} \textbf{ otherwise} \end{cases} \tag{1. 7}$$

(Qin and Tang, 2013)

The ROC curve is produced by letting T vary from $0 \ \to 1$. In other words, each dot in the ROC curve represent a new decision threshold to the classifier and the full ROC curve consists of the TPR and FPR for each decision threshold $T \in [0,1]$. In figure 7 below, an example of a ROC curve is illustrated.

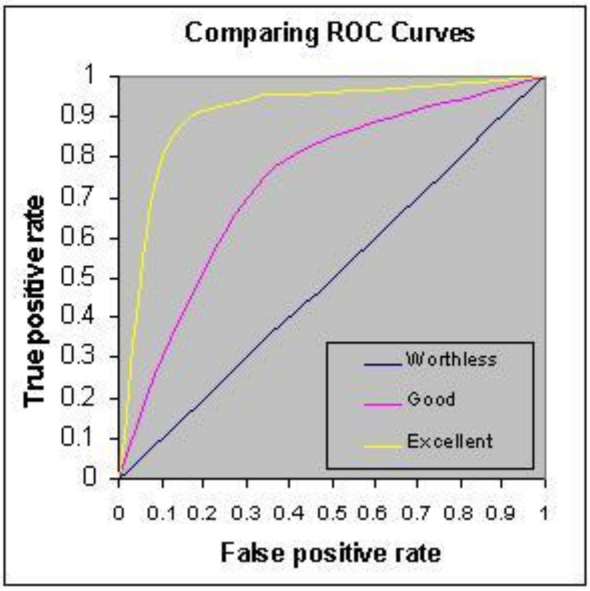

Figure 7: ROC curve examples (Tape, 2018)

Since the ROC curve does not only provide information on the accuracy of the classifier, but also gives a relation between TPR and FPR it makes it possible to judge a classifier when the cost of error is different depending on if the error is missing a member of the positive class or a member of the negative class. (Qin and Tang, 2013)

2.3.5.3.1 Area Under the ROC Curve

As concluded in the previous section, a ROC curve gives a graphical interpretation of how well a classifier performs. The ROC for a perfect classifier would have 100% true positive rate with 0% false positive rate for all T and the ROC curve would therefor go from (0,0) to (0,1) to (1,1). A classifier that is randomly guessing whether an observation belongs to a certain class would produce a ROC curve that goes from (0,0) to (1,1) and a classifier that gets each classification wrong would have a ROC curve that goes from (0,0) to (1,0) to (1,1). To get a single value that represents this, the area under the ROC curve (AUC) is calculated. (Qin and Tang, 2013) The AUC for a classifier that is simply guessing would thus be 0.5, while AUC for a perfect classifier is 1.0.

It has been proven that this way of measuring performance is both more consistent and discriminating than the accuracy measurement. (Ling, Huang and Zhang, 2003)

2.4 Features

In the vast number of previous papers that have predicted churn among telco companies and within other industries (Hung, Yen and Wang, 2006; Jahromi *et al.*, 2010; Portela and Menezes, 2010; Zhang *et al.*, 2012; Kirui *et al.*, 2013; Backiel, Baesens and Claeskens, 2014; Kim, Jun and Lee, 2014; Huang *et al.*, 2015; Kim, Kim and Park, 2017; Subramanya and Somani, 2017; Mitrović *et al.*, 2018), many different features have been identified as effective when identifying churn. Many of these will most probably be applicable on Rebtel's customers and therefore previous churn prediction papers will be one of the primary sources for features in this paper. Many of the features that previous papers have used in their classification models are not possible to use since the data is missing, an example of such data is demographic data. Due to the particularities of the business of Rebtel, some features do not come from previous academic work, but are rather derived through discussions with commercial and product usage analysts at the firm. These features are based on previous internal analysis done within Rebtel.

The features are divided into three categories; *Activity Features, Customer Profile Features, Payment Features*.

2.4.1 Activity Features

Activity features aim to describe the customers' behavior and usage of the services provided by Rebtel. For churn in telco companies, activity features includes both outgoing activities and ingoing activities (e.g. calls, text messages) (Hung, Yen and Wang, 2006; Kirui *et al.*, 2013; Huang *et al.*, 2015) but in the case of Rebtel there is only data on the outgoing activity. Within software products it has been shown that both recency and frequency of activity measures are influential when predicting churn (Castro and Tsuzuki, 2015) and therefore both time invariant and time variant features are included.

Many of Rebtel's subscription customers regularly deactivate and reactivate their subscription during the timeframe when the subscription is active. The exact reason for this is not confirmed, but one hypothesis is that the customer wants more control over their subscription and does not want it to renew automatically. This event would of course be a good activity feature to include in the model, but since the logging of such events started mid-November, not enough data is available to include it in the model.

2.4.1.1 Time invariant features

The activity features that have been identified in previous papers and together with experts at Rebtel that are invariant to time are presented below in table 1. All are measured in the 30 days period prior to the evaluation date if nothing else is mentioned.

Table 1: Time invariant activity features

Feature	Variable name	Source
Number of calls	CountCalls	(Hung, Yen and Wang, 2006; Jahromi *et al.*, 2010; Zhang *et al.*, 2012; Kirui *et al.*, 2013; Huang *et al.*, 2015)
Total call duration	TotalDuration	(Kirui *et al.*, 2013; Huang *et al.*, 2015)
Number of days at least one call has been made	DaysCallsMade	(Jahromi *et al.*, 2010; Zhang *et al.*, 2012; Kirui *et al.*, 2013)
Number of calls < 30 seconds	NbrShortCalls	(Subramanya and Somani, 2017)
Proportion of calls < 30 seconds	PropShortCalls	(Subramanya and Somani, 2017)
Number of failed calls	FailedCalls	(Subramanya and Somani, 2017)
Proportion of failed calls	PropFailedCalls	(Subramanya and Somani, 2017)
Number of different numbers called (B-numbers)	NumberOfBNumbers	(Kim, Jun and Lee, 2014)
Time since last call	DaysSinceLastCall	(Jahromi *et al.*, 2010)
Time between two latest calls	DaysDiffBetweenCalls	Rebtel analysts.
Number of contacts in phonebook from the main destination	NbrOfContacts	Rebtel analysts.

2.4.1.2 Time variant features

To record changes in behavior over time, trends for some of the main activity features are also added as features. In table 2 below, these are presented. The trends are recorded over both 30 days and 15 days in the case of number of calls and call duration.

Table 2: Time variant activity features

Feature	***Variable name***	***Source***
Trend in number of calls over 30 day period	CallTrend30	(Zhang *et al.*, 2012; Mitrović *et al.*, 2018)
Trend in number of calls over 15 day period	CallTrend15	Rebtel analysts.
Trend in call duration over 30 day period	DurationTrend30	(Kirui *et al.*, 2013)

Trend in call duration over 15 day period	DurationTrend15	Rebtel analysts.
Trend in number of different numbers called (B-numbers) over 30 days period	BNumbersTrend30	Rebtel analysts.

2.4.2 Customer Profile Features

Customer profile features play a substantial role for churn predictions in traditional telco companies (Backiel, Baesens and Claeskens, 2014; Kim, Jun and Lee, 2014; Subramanya and Somani, 2017), but due to the difficulties of getting hold of demographic data in many tech industries there are often fewer demographic data in software-based company churn predictions (Castro and Tsuzuki, 2015; Periáñez *et al.*, 2016). In industries where demographic data on the customers is hard or impossible to obtain, customer profile features are based on variables such as tenure with the service (Periáñez *et al.*, 2016), country of origin and device type (Sifa *et al.*, 2015). Since Rebtel does not have demographic data on their customers, the features describing customer profile share more resemblance with papers investigating churn in mobile application products. All the customer profile features are presented in table 3 below.

Table 3: Customer profile attributes

Feature	***Variable name***	***Source***
Tenure	MonthsSinceSignup	(Jahromi *et al.*, 2010; Zhang *et al.*, 2012; Backiel, Baesens and Claeskens, 2014; Periáñez *et al.*, 2016; Kim, Kim and Park, 2017; Subramanya and Somani, 2017)
Device type	Device	(Kim, Jun and Lee, 2014; Sifa *et al.*, 2015; Mitrović *et al.*, 2018)
Origination country	Origin	(Backiel, Baesens and Claeskens, 2014; Huang *et al.*, 2015; Sifa *et al.*, 2015; Subramanya and Somani, 2017)

2.4.3 Payment Features

Several studies have included payment features when predicting churn, in both telco settings and other (Backiel, Baesens and Claeskens, 2014; Periáñez *et al.*, 2016; Mitrović *et al.*, 2018). In the case of Rebtel this might be interesting since the payment process looks a little different depending on which device and product the customer is using. Android users for instance can set up automated payments towards a monthly subscription in the app, while iOS users are limited to Apple’s in-app-purchase if paying in the app, which are both more expensive and cannot be automated. An Apple user can although sign up for automated payments through the web interface, so this difference would not be picked up on solely by looking at the device that the user is on. A study on churn in a Singapore telco company found that churn differed a lot between different payment methods and hypothesized that this had to do with the difference in actively having to stop paying (with automated payments) or simply not paying (Au, Chan and Yao, 2003). Other payment feature used in previous academic works include billing amount and number of payments. In table 4 below all payment features are presented.

Table 4: Payment features

Feature	Variable name	Source
Number of payments	NbrPayment	(Backiel, Baesens and Claeskens, 2014; Kim, Jun and Lee, 2014; Subramanya and Somani, 2017; Mitrović *et al.*, 2018)
Number of denied payments	DeniedPayments	(Hung, Yen and Wang, 2006)
Total billing amount	SumTotBilling	(Hung, Yen and Wang, 2006; Jahromi *et al.*, 2010; Zhang *et al.*, 2012; Huang *et al.*, 2015; Subramanya and Somani, 2017)
Days since latest payment	DaysSinceLatestPayment	(Backiel, Baesens and Claeskens, 2014)
Most common payment method	PaymentMethod	Rebtel analysts.

2.5 Feature importance

As mentioned in the section *Random Forest*, the feature importance can be derived through the random forest algorithm by measuring the mean decrease in Gini impurity in a node using a certain feature and averaging it over the number of observations ending up in a specific node. This section aims to explain this process in more detail and how the resulting feature importance should be interpreted.

2.5.1 Algorithm

To derive the importance of a feature X_m for predicting Y, weighted mean Gini impurity decrease is summarized for all nodes t where X_m is used. Consider a random forest with N_t trees, then:

$$Imp(X_m) = \frac{1}{N_T}\sum_{T}\sum_{t\in T:v(s_t)=X_m} p(t)\Delta i(s_t, t)$$

Where each tree $t \in T$, each split is denoted s_t, the decrease of Gini impurity in a certain split is $\Delta i(s_t, t)$, $p(t) = \frac{N_t}{N}$ is the proportion of samples reaching t and $v(s_t)$ is the variable used to in split s_t. (Breiman, 2001)

It has been shown that this way of measuring feature importance performs well when investigating what effect a certain feature has on the classification of the observations and that the feature importance is zero if and only if the feature is irrelevant. (Louppe *et al.*, 2013)

2.5.2 Interpretation of Feature Importance

The mean Gini impurity decrease gives a good estimation of how important a feature is by measuring which features that are more closely related with the dependent variable. But it does not say anything about the relationship between the independent variable and the dependent variable. Therefore, feature importance is only used to identify *which* features that impact the churn prediction to a high degree.

2.5.3 Partial Dependence Plots

When the most influential features have been identified through their feature importance, the next step is to investigate how they influence the target variable. Hastie, Tibshirani and Friedman, (2009) suggest doing this through partial dependence plots.

This is done by fixing the identified feature, X_m, to different values for each sample in the training data and calculating the average probability output of the model for each value of X_m. This will give the probability of an observation being a churner or not, conditioned on a certain value of X_m and will be used to interpret the effect of a certain feature on the outcome.

The values of the partial dependence are plotted and show the average change in probability of churning when X_m takes different values.

2.6 Summary of Theoretical Framework

Previous research have made known that a high customer retention is vital for both financial and user base implication. It has also been shown that prediction techniques such as the Random Forest algorithm are successful for churn prediction in other settings, that share resemblance with an app-delivered service for international calling. The Random Forest algorithm also provides a feature importance which can be used to further investigate drivers of churn.

All these learnings are used when developing the method employed in this paper, which is described in the following chapter.

3 Method

This section aims to explain the method employed in this paper. The characteristics of the research approach are explained, and the individual components are explained in greater depth. A method outline and summary is also presented.

3.1 Research characteristics

Depending on the goal and the characteristics of the research, the methodology of the paper can be either descriptive, explanatory, exploratory or problem-solving (Höst, Regnell and Runesson, 2007). Since the objectives of this paper are dual, the research approach can also be seen as a combination of being both problem-solving and explanatory. To answer the research question of whether it is possible to identify individual customers that have a high propensity to churn, a problem-solving approach has been employed. When investigating which customer features that have a high impact on the churn probability, a more explanatory approach has been employed.

In order to map and understand previous academic work in relevant areas a literature review has been also been performed. This was performed in an exploratory manner by identifying key words in the relevant fields of studies and performing literature searches through search engines that gather previous academic research. Through the initial findings more relevant key words were identified and the literature review continued with these in an exploratory, iterative manner.

3.2 Literature review

The aim of the literature review is to gain understanding of previous research in the areas of customer retention and churn prediction. It is also the fundament of deriving features to include in the modelling and how the results of this study relate to previous findings on customer loyalty in the medium and long-term phase of the customer relationship lifecycle.

The literature review is conducted in an exploratory manner where relevant keywords are used to perform literature searches, mainly through library catalogues, the Elsevier ScienceDirect database and through Google Scholar. New keywords are later discovered in the relevant literature and the process is therefore iterative in its nature.

The literature identified through the exploratory research is also complemented by advices from University supervisor, Rebtel analysts and industry reports purchased by Rebtel.

3.3 Data

Based on previous papers on churn prediction and insights from Rebtel analysts the features described in the *2.4 Features* section are chosen as candidates for variables to be used in the churn classification model. In order for these to be tested in a model raw data has to be extracted from Rebtel's databases and turned into features. This section aims to describe this part of the method.

3.3.1 Retrieval, Feature Extraction and Preprocessing

Rebtel, as most modern companies, stores extensive amounts of data on their customers. This data is fetched from different operating systems and stored in a database that is referred to as the data warehouse. Data from the data warehouse can later be queried by different reporting tools or through SQL queries directly into the database.

As a first step in constructing features, raw data is queried and, if necessary, preprocessing the raw data into a workable format. As an example, calling data is stored with a new row for each call attempt made by a user. In order to turn this data into a feature such as *call duration*, the duration for each successful call attempt for a certain customer must be summed over a specified timeframe. If the trend of *call duration* is to be used as feature, the same must be done for data over several sequential timeframes and then have the trend calculated over these.

Preprocessing might also include handling observations with missing values, transforming data into categorical data and binning data.(Kotsiantis, Kanellopoulos and Pintelas, 2006)

Some features in this paper will be retrieved directly from the data warehouse, such as *origin* and *device.* While others will need some preprocessing in order to extract the features. The trend features will be calculated using the value of two consecutive time periods, t_0 and t_1, by calculating the relative change:

$$trend(x) = \frac{x_{t_1} - x_{t_0}}{x_{t_0}}$$

Meaning that trend values will range between -1 and infinity. This follows the methodology employed by Mitrović *et al.* (2018).

Categorical variables with many different categories and skewed distributions will also be binned. An example of such a feature is origin country of which there are about 100 categories, but more than 90 % of the observations belong have origin that is one of United States, Canada, New Zeeland or Singapore. In this case the categories will be one of the top four countries or “Other”.

3.3.1.1 Segmentation

As the aim of this paper is to investigate medium and long-term retention factors for customers calling to India, not all customer data can be used in the modelling. Also, we do not want to include dormant users in the modelling.

First, a reminder of the difference between a churned user and a dormant user. As defined in section *2.3.3 Churn and Inactivity* a churned customer is someone who has been a customer but currently has no likelihood of undertaking new purchases. A dormant user is someone who, considering their usual buying pattern, shows a weak probability of repeating a purchase in the near future.

To make this distinction in practice and to exclude both users that are new to the product and users that can be assumed to be dormant, some users are excluded from the segmentation used in the modeling. Only users that placed at least one call, with an India subscription product, in each of the two previous 30 day periods prior to the evaluation date are included. This will minimize the number of dormant users included and make sure that no new users are present in the dataset used for modeling.

3.4 Modelling

As the modeling of churn prediction is a crucial part of this paper, this section will go through the model choices that have been made and explain how the modelling is done. All modelling is done using the k-fold stratified cross-validation approach described in section *2.3.4 K-Fold Stratified Cross-Validation* to make sure that the model evaluation is performed correctly.

3.4.1 Initial Model and Iterations

As a first step, the features described in section *2.4 Features* are added to an initial model which is trained on randomly under sampled data according to *Figure 6*. Although the random forest is robust to noise and irrelevant variables, the expected inclusion of some redundant variables in the initial model will have a limited negative effect on the model metrics. But, since the feature importance is zero if and only if the feature in question is irrelevant (Louppe *et al.*, 2013), the initial model can be improved by removing any such irrelevant features and thus noise from the input variables.

The algorithm used for the classification is a random forest with 1000 trees.

3.5 Measuring Importance and Impact of Features

When a satisfactory model is achieved, the feature importance and impact of features is derived. The feature importance is calculated by weighted mean Gini impurity decrease as described in section *2.5.1 Algorithm* and the impact of features with high importance are further investigated by plotting partial dependence plots as described in section *2.5.3 Partial Dependence Plots*.

3.6 Validation of Findings and Model Performance

To ensure that the results of the final model regarding predictive power and performance, the results are validated through comparing with previous academic papers on churn prediction. The final model is compared on AUC score with churn prediction models from previous research to ensure that the model is satisfactory.

3.7 Method Outline and Summary

This paper aims to answer three questions, one regarding what the drivers of churn are, a separate question is how churn can be predicted and the final question is regarding how Rebtel should use the learnings from the previous two questions to improve their operations. Although different, the approaches to answering these questions are intertwined and the answers to what the drivers of churn are, heavily depends on being able to successfully predict which users that are going to churn. The general steps employed in this paper are explained below.

1. **Literature review**
   A literature review, investigating and compiling previous academic research within customer retention, churn prediction, relevant customer features and attributes, and churn prediction modelling is performed. The findings of this part are presented in *2 Theoretical Framework*.
2. **Data collection**
   Data is retrieved from Rebtel databases and relevant attributes and features that have been identified through discussions with Rebtel analysts and through the literature study are engineered or extracted from the raw data. Data retrieval is done through SQL queries while the extraction and feature engineering are done both within the Rebtel database infrastructure, using SQL querying, and on local machines using Python.
3. **Developing a model to predict churn**
   A churn prediction model is developed. This is done using features from previous similar studies and combining those with features believed by Rebtel analysts to be predictive of churn. A random forest classifier is trained and evaluated using the K-Fold Cross-Validation technique described in section 2.3.4 to ensure that the model is satisfactory.
4. **Validate model performance**
   The final model is benchmarked against models from previous academic papers to ensure that the performance is satisfactory.
5. **Identify drivers of churn**
   Using the churn prediction models from step one, features with high importance will be extracted and investigated by creating partial dependence plots. These plots will give an indication of what the dependency between the feature and the churn variable is.
6. **Analysis and recommendations**
   Use the learnings from the above points to derive concrete recommendations for future improvement in the customer retention area for companies in the software-based international calling industry.

Steps 2, 3 and 5 are depicted in *Figure 8* below, which shows the modelling and extraction of feature importance and effect. This sub-process follows the conventional machine learning methodology (Sarkar, Bali and Sharma, 2018).

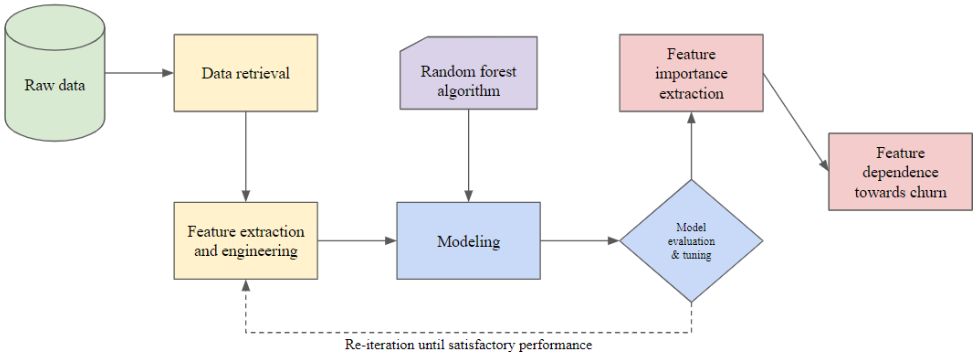

Figure 8: Methodology for developing a model to predict churn.

In the following chapter the empirical findings of the churn prediction model development, validation of the model performance and identification of features that drive churn are presented.

4 Empirical Findings

4.1 Data Characteristics

The dataset used for modeling contains 96 026 observations of customers that were active at the observation date, 2018-09-01. Out of these 10.95% churn in the coming month.

38% of the customers are based in the United States, 31% in Canada and the 11% in New Zealand, the rest are scattered over the world. The majority of customers have signed up through Rebtel's Web portal or through the Android app, while 29% have signed up through an iOS device.

A more detailed description of the numeric features in the dataset can be found in *Appendix - 1.1 Data description*.

4.2 Model performance

The final model was trained and evaluated using the K-Fold Cross-Validation technique described in section *2.3.4 K-Fold Stratified Cross-Validation*. In the iterative model building depicted in *Figure 8* it was found that under-sampling did not improve the performance of the model and therefore the final model does not include this step.

As described in section *2.3.5 Evaluating Model Performance*, AUC is one of the most consistent ways of evaluating the performance of a classifier such as the one employed in this paper and this is the main evaluation used in the modeling process. The ROC curve for the final classifier evaluated with a 6-Fold Cross-Validation technique is depicted in *Figure 9* below.

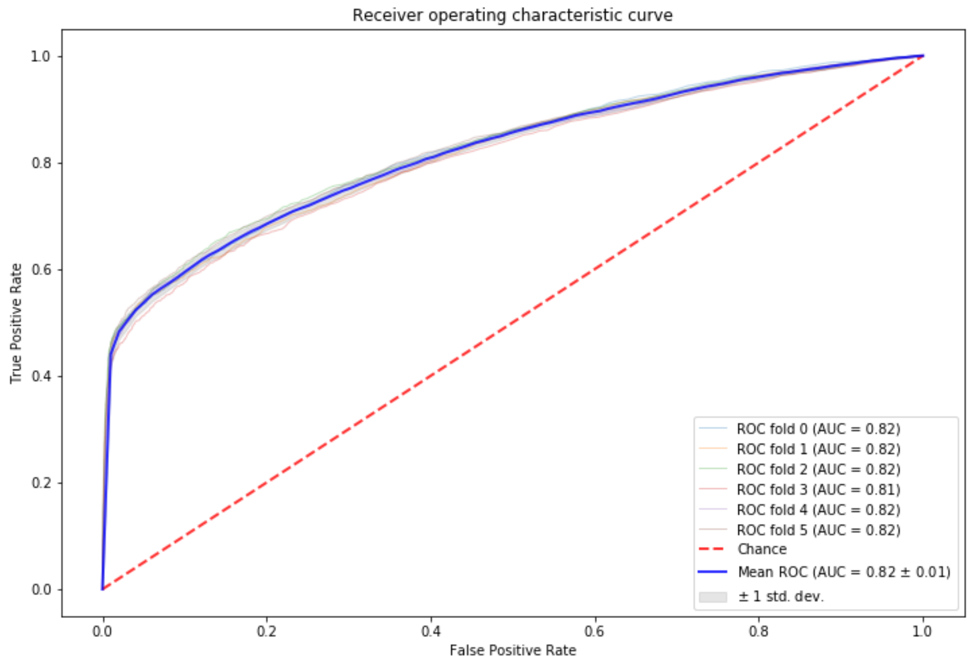

Figure 9: ROC curves for the final model.

The average AUC for of the folds is 0,82 with a standard deviation of 0,01. Other evaluation metrics for each of the folds are presented in the table below.

Table 5: Performance metrics of final model

Metric	*Fold 0*	*Fold 1*	*Fold 2*	*Fold 3*	*Fold 4*	*Fold 5*	*Average*
Accuracy	94,85%	94,45%	94,12%	93,23%	91,96%	89,57%	**93,03%**
Recall	43,99%	44,80%	43,92%	42,46%	44,86%	47,22%	**44,54%**
Precision	82,17%	82,04%	85,41%	81,88%	82,62%	87,64%	**83,63%**
F1-score	0,5731	0,5795	0,5801	0,5592	0,5815	0,6137	**0,58**

4.3 Features

In this section the features role in the final model is described. First, the importance of the features when determining if a customer has churned or not is presented and then a deeper analysis of how the most important features affect if the customer churns or not is visualized through partial dependence plots.

4.3.1 Feature importance

The final model is used to calculate the feature importance according to the algorithm described in section *2.5.1 Algorithm*. The resulting feature importance is presented in the figure below.

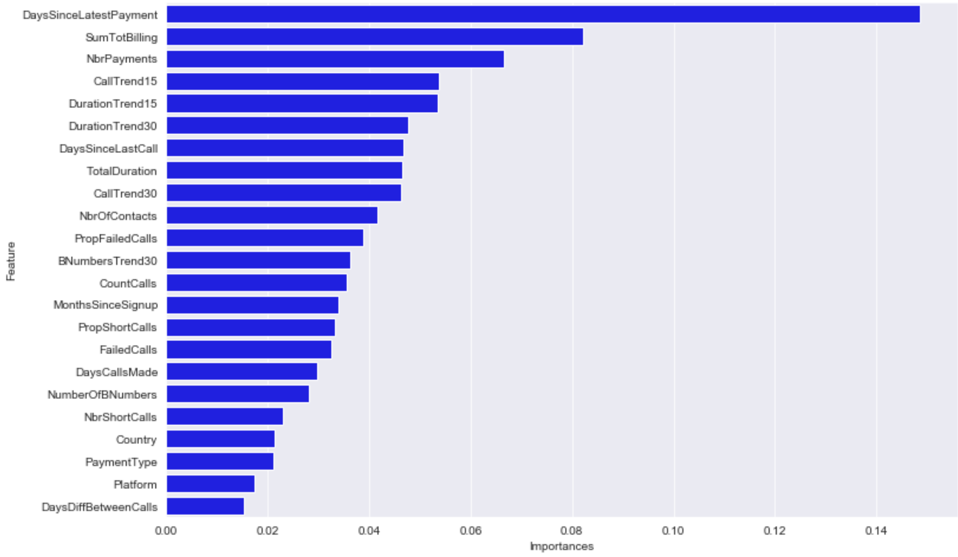

Figure 10: Feature importance

4.3.2 Partial Dependence Plots

To determine the effect of a specific feature on the predicted outcome, partial dependence plots are employed. This follows the method described in *2.5.3 Partial Dependence Plots*, in each plot both the mean change in predicted outcome and a number of individual churn predictions are plotted. The thick blue and yellow line represents the average change in predicted probability of churning in comparison to the leftmost point on the x-axis. In other words, a y-value of -0.1 represent a 10%-unit decrease in probability of churning compared to the initial, leftmost, value on the x-axis.

4.3.2.1 Days Since Latest Payment

For the most important feature, days since latest payment, the partial dependence plot is shown below.

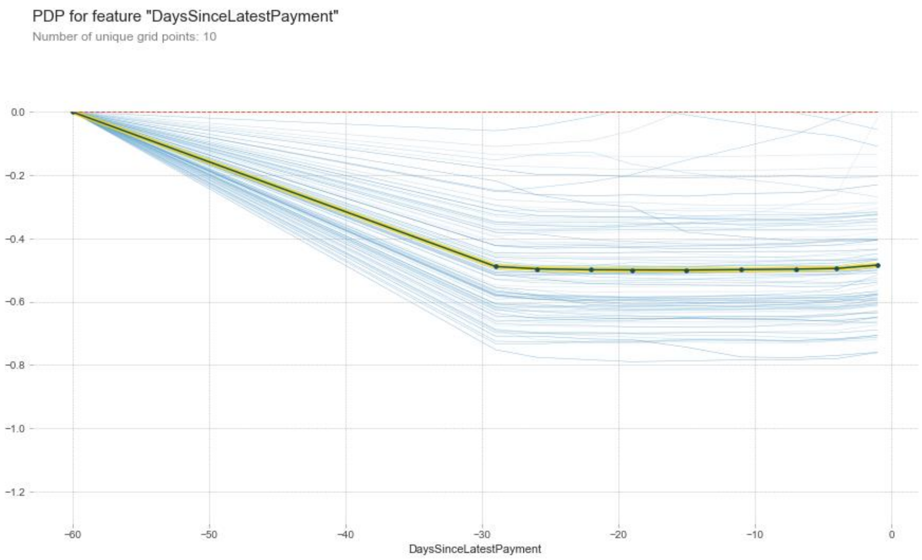

Figure 11: Partial dependence plot for days since latest payment

As can be seen in *Figure 11* the probability of a customer to churn if the latest payment was 60 days ago is about 50%-units higher than for a customer that made their latest payment 28 days ago. Other payment variables such as total payment last 30 days (SumTotBilling) and number of payments in the last 30 (NbrPayments) show similar patterns – if a payment has been done in the last 30 days then the risk of churning becomes lower.

4.3.2.2 Call trend in the last 15 days

The most important activity feature was call trend in the last 15 days. The effect of this feature on the predicted churn probability is shown below.

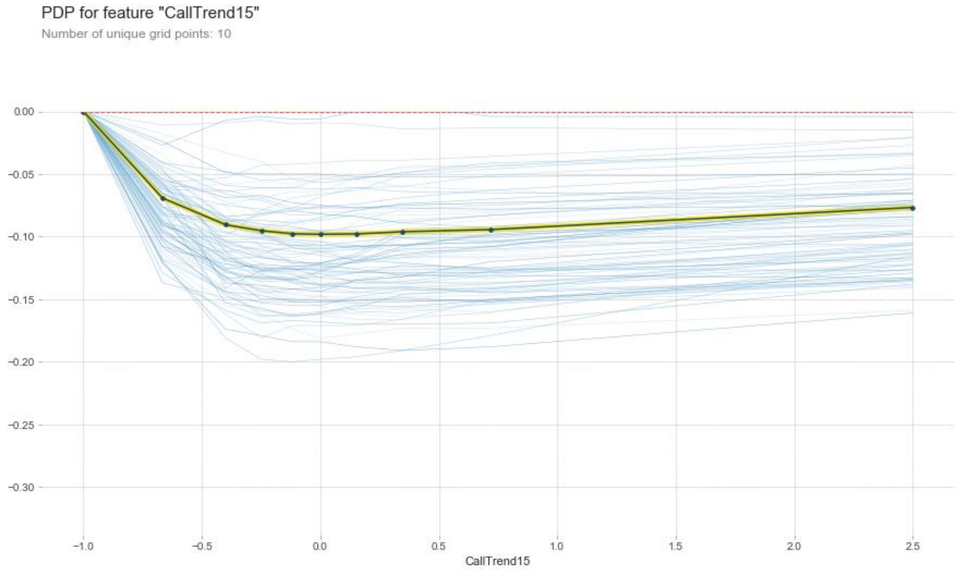

Figure 12: Partial dependence plot for call trend in the last 15 days

4.3.2.3 Payment method

Although not one of the more important features, the partial dependency plot for payment methods show that how the customer pays plays a role in determining the probability of that customer churning. If a customer used Apple's in-app-purchase (PaymentMethod_API) to purchase the product, that customer is approximately 10% units more likely to churn than for customer using other payment methods. These are the users using Apple's in-app-purchase which is described in section *2.4.3 Payment Features*. This is illustrated in the partial dependence plot below.

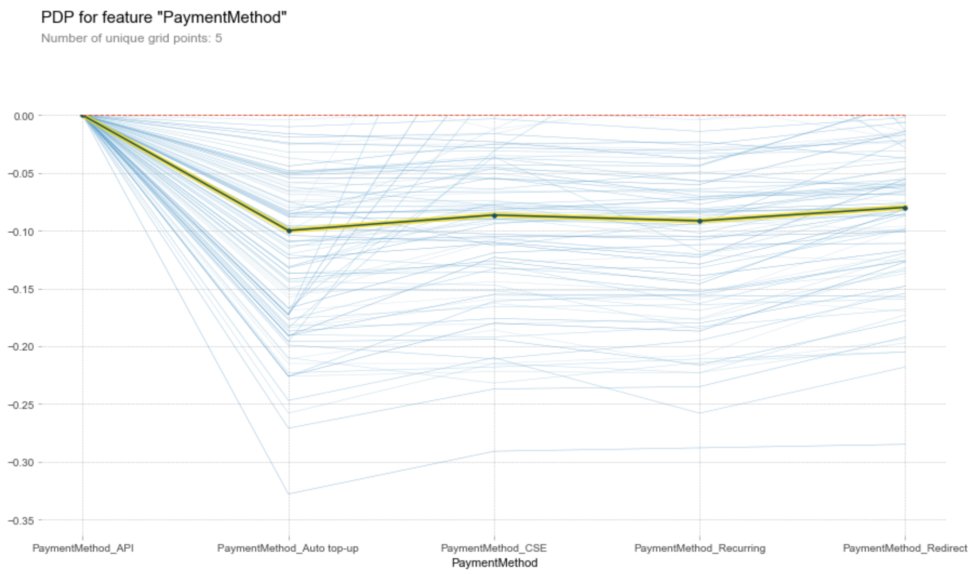

Figure 13: Partial dependence plot for payment methods.

5 Analysis and Discussion

In this chapter, the final model and feature effects on churn are analyzed and discussed. The analysis and discussions are centered around the three main research questions; how churn can be predicted, what the drivers of churn are and what can be done to improve the medium phase and long-term retention at Rebtel.

5.1 Predicting Churn

As one of the main objectives of this paper is to be able to identify which customers that are more likely to churn, this section goes through the results and compares them to previous papers to check if the predictive power is at a good enough level.

5.1.1 Comparison of AUC

As described in section *2.3.5 Evaluating Model Performance* there are many different ways of measuring the performance of a classifier. In this paper the focus has been on AUC since it gives a more consistent and discriminating picture of the performance than measuring accuracy and to avoid the accuracy paradox. To determine whether the model in this paper produces prediction with sufficient predictive power the AUC is compared to recent similar studies in both the telecommunications industry and the Mobile Games industry. A table of the comparisons is presented below.

Industry	***Paper***	***AUC***
Telecommunications	(De Caigny, Coussement and W. De Bock, 2018)	0,667
Telecommunications	(Mitrović *et al.*, 2018)	0,735
Telecommunications	(Zhu, Baesens and vanden Broucke, 2017)	0,805
Telecommunications	(Hassouna *et al.*, 2015)	0,763
Mobile Games Application	(Castro and Tsuzuki, 2015)	0,790
Mobile Games Application	(Periáñez *et al.*, 2016)	0,960
Mobile Games Application	(Milošević, Živić and Andjelković, 2017)	0,83
Telecommunications/ Mobile Application	This paper	**0,82**

In general, the model performs slightly better than its' peers within telecommunications. This indicates that the behavioral patterns for Rebtel customers are more predictable than for an average telecommunications company. One reason for this might be that in the Rebtel case, more relevant data is available such as access to the customers' saved contacts. The fact that this product is only used to call international calls to a certain destination can also be a driver of higher similarities between individuals. Another reason can be that the internal definition of churn used at Rebtel makes it easier for a model to predict if a customer will be deemed churned in the future, which is further discussed in section *5.1.3 Impact of Churn Definition*.

Although, if comparing the AUC to other mobile applications such as mobile games, the AUC is rather low. The reason for comparing the AUC of this paper to prediction models within mobile games applications, is that it is the most similar area, other than pure telecommunications, that have appeared in academia in recent years. It might be that customer behavior for an application such as Rebtel's lies in-between conventional telecommunications services and app delivered mobile games when it comes to churn, which would explain the model's performance that lies between the two sectors when it comes to recent previous papers.

Since the AUC measurement is an indicator of the TPR in comparison to the FPR, it can be concluded that the predictive power is on par with previous papers in similar industries.

5.1.2 Impact of Churn Definition

In this paper Rebtel's internal definition of churn has been applied. This states that a customer is considered churned 30 days after the last call or payment. Furthermore, a monthly payment is considered as 30 small, daily payments as revenue is recognized during the validity of the product.

This definition becomes evident in the model since customers who made their latest payment 59 days ago, will churn the coming day if no payment is made. This explains why the days since the latest payment becomes the feature with the highest importance (*Figure 10*), since it can with a high degree of certainty separate customers that will churn by choosing everyone that hasn't performed a payment in about 59 days. This is also the most probable reason for total billing amount in the last 30 days and number of payments in the last 30 days become the second and third most important customer features, since it can be assumed the collinearity between these and days since latest payment is high.

The fact that the model has an easy time separating some of the churners is also evident in the ROC curve (*Figure 9*). In the beginning, the curve is very steep which means that some of the observations included in the sample are obvious churners and the model manages to identify these while keeping the FPR close to zero.

The churn definition applied in this paper builds in unnecessary amount of time between the last action performed and the churn event. A different definition would probably make it possible to decrease the time it would take to identify churners and would therefore most probably enable earlier identification of churners.

5.2 Future Use of Churn Prediction at Rebtel

This chapter investigates how the results from the churn prediction can be used in a business setting to decrease churn and thereby improve the life time and life time value of the customers.

5.2.1 Recall and Precision

As explained in section *2.3.5.2 Recall, Precision, F1 score and the Accuracy Paradox* the accuracy is not a suitable way to measure the performance of a model that classifies an imbalanced dataset, due to the accuracy fallacy. Two measurements that on the other hand tells us a lot of how the model manages to identify churners and non-churners in the test set are recall and precision.

As a reminder from the theoretical framework, recall tells us how big part of the customers who actually churn the model manages to correctly identify as churners. While precision measures the proportion of actual churners among the identified churners.

These two measurements have a great impact on what measures that can be employed and the efficiency of these if the results of the prediction is used to decrease churn. This can, for instance, be by offering a promotional price or other incentives to stay as customer. Such measures can never decrease churn more than at the level of precision of the model and since all identified churners that take up such an offer will impose a cost for the firm, the recall will have a big impact on the cost efficiency of such measures.

The average recall of our model is 44.54% which means that almost half of all churning users can be reached with incentives to stay. The precision of 83.63% indicates that less than 17% of the customers that are offered any incentives are done so “in vain” since they would have stayed as customers anyway.

When designing any churn prevention measurements with the churn prediction in this paper as basis, the precision and recall will be instrumental when deciding which financial costs that can be associated with the measurements taken.

5.2.2 Measures to Keep Customers Identified as Churners

Several previous studies have tested what measurement that can be used to incentivize customers that are identified through a churn prediction model to stay. In a study investigating which measures that reduce churn in a similar setup as in this case, the authors found that giving free incentives, organizing special customer events and obtaining feedback through questionnaires all gave a significant reduction in churn for the customers identified as potential churners. (Burez and Van den Poel, 2007) The phenomenon of churn decreasing after being presented a satisfaction questionnaire is known as the “mere-measurement effect”. (Morwitz and Fitzsimons, 2004)

Of course, each different retention action comes with different amounts of costs and the cost efficiency depends on both how many customers that are retained and the cost associated to the action.

In the case of Rebtel, many different approaches should be tested and evaluated to see which actions that retain customers in an efficient way. Sending satisfaction questionnaires has the benefit of also providing the firm with data on what the actual reasons for churn are and can be of help in future product development.

Ideally, any measures taken to deter customers from churning should also address the underlying reasons for why the customers wanted to leave in the first place.

5.3 Drivers of Churn

The second research question of this thesis, regarding which the drivers of churn during the medium phase and long-term retention period are, were investigated by plotting partial dependence plots for some of the features used by model. When looking at the features one by one, the general effects of each feature becomes rather self-explanatory and obvious. These measured effects include,

- Customers who have made a payment in the last 30 days are less likely to churn than customers who have not. See *Figure 11*.
- Customers who have called substantially less in the last 15 days compared to the previous 15-day period are more likely to churn. See *Figure 12*.
- Customers who use Apple's in-app-purchase and subsequently pay more and cannot activate automatic payments, are more inclined to churn than customers who use other payment methods. See *Figure 13*.

These conclusions were not surprising for managers and analysts at Rebtel and the effects were already known. The findings do neither point out many clear areas of improvements when it comes to product development since features such as call trend and days since payment rather demonstrates behavior that comes as a consequence of the root cause of the customer's dissatisfaction with the service. Although, one product improvement that is clear is that if the payment options for Apple users would be exchanged for the same payment options as for Android, the churn would decrease in this group. The model indicates that churn would decrease by approximately 10% units for each Apple customer if this was applied (See *Figure 13*).

The partial dependence plots together with the feature importance fill an important function when it comes to explaining an otherwise inexplicable model which is crucial for taking business decisions based on the outcome of the churn prediction model.

6 Recommendations and Conclusion

The final part of this paper is to conclude recommendations for Rebtel on what should be done to improve the medium phase and long-term retention for customers calling to India. This section concretizes the findings in this paper into recommendations for future actions and summarizes the findings in general. It also gives suggestions for future research.

The aim of this paper has been to build a model that identifies customers that have a high probability of churning while also determining some key drivers of churn. This paper has shown how a model can be built that identifies customer that have a high probability of churning but the model fails to give enough insights into what the key drivers of churn are, other than on a basic level.

6.1 Recommendations

Rebtel is acting in a market that is in decline and one of the most important assets the company has is the customer base that it has built up so far. To continue the success in a market that is in decline, the firm should focus on decreasing the churn.

For customers calling to India, potential churners should be identified through a Random Forest classification such as the one developed in this paper and these should be incentivized to stay as customers. What measures that should be used to incentivize user to stay has not been investigated in this paper, and several options should be tested. When designing these retention actions, the recall and precision of the model should be considered as these will impact the cost efficiency of the measures taken. For other customers such as subscription customers calling to other destinations than India and pay-as-you-go customers, similar models should be developed in order to prevent users to churn in these markets as well. If a simpler, rule-based, model is developed for any of the other customers groups, the model in this paper indicates that payment features are the most relevant when it comes to predicting churn and these should therefore be included.

Product wise, Rebtel should explore the possibilities to replace the in-app-purchase in Apple iOS with the same payment options available for Android users. A successful implementation of this would lead to a decrease of churn with 10% for the customer group that are using Apple's in-app-purchase today.

6.2 Contributions

The main contribution of this paper is the application of a churn prediction model specifically for the $70 billion (Telegeography, 2017) international calling market, something that has not been found in previous research. It is also the first study to investigate churn in a service that lies in-between a well-researched traditional telco industry and a modern, quickly evolving tech industry.

A lot of inspiration has been gathered from previous studies investigating churn prediction within the telco industry, especially when it comes to features used in the classification, some of the features that are brought forward in this paper are newly derived thanks to the knowledge among Rebtel analysts. These features are:

- Time between two latest calls
- Number of contacts in phonebook from main calling destination
- Trend in number of distinct numbers being called
- Most common payment method on different mobile devices

These are features that can be emulated to be used in both ordinary telco settings and in eventual future research within the international calling market.

Finally, this paper does an attempt at using a complex machine learning algorithm to provide general learnings in what the drivers of churn are. Although the learnings in this area are limited, this is something which has not been attempted in previous churn prediction studies.

6.3 Future Research

Although this thesis manages to identify users that are about to churn and give recommendations on the future use of these findings, it does not manage to answer the question of what the underlying drivers of churn are. These studies are often tightly tied to the industry they are investigating and for the area of the international calling market no specific study has been found.

In this area other approaches than Random Forest should be investigated and compared with the findings of this paper. Future research should also identify *early* indicators of churn as done within other industries (Khan *et al.*, 2015) as an earlier detection of churn would give more room to convince the customer to stay.

When giving recommendations on how to implement the churn prediction in a business setting, it also became clear that the research within this area is insufficient. None of the previous research that have been investigated relate the performance of the churn prediction to business implications. The financial and operational implications of metrics such as recall, precision, F1 score, AUC and accuracy is an area where research is needed.

6.4 Conclusion

This paper has shown that it is possible to accurately predict which individual users that are about to churn for an app-delivered international calling service. The performance of the prediction is slightly higher than several similar studies made with pure telecommunications datasets and slightly worse than many previous studies investigating churn in mobile games applications. The features used in the modelling is a mix between features that are specific for an app-delivered international calling service and general telecommunications features which have been used in previous churn prediction papers.

Isolating the features one by one through feature importance and partial dependence plots gives an indication of which features that are important to determine which customers that are likely to churn and how each feature affect the probability of a customer churning. But they give limited insights into what the drivers of churn are and should rather be used as a baseline for future research into the actual reasons for why customers churn.

Machine learning algorithms are well suited to accurately predict which customers that will churn, by detecting complex user feature patterns that indicate churn. But the learnings when isolating features one by one are limited and when it comes to investigating the underlying causes to why customers churn, they become banal. The primary use for these should be to validate the model and understanding how it works.

7 References

Atkins, C., Gupta, S. and Roche, P. (2018) *Introducing customer success 2.0: The new growth engine*, *McKinsey&Company*.

Au, W. H., Chan, C. C. and Yao, X. (2003) 'A novel evolutionary data mining algorithm with applications to churn prediction', *IEEE Transactions on Evolutionary Computation*. doi: 10.1109/TEVC.2003.819264.

Backiel, A., Baesens, B. and Claeskens, G. (2014) 'Mining telecommunication networks to enhance customer lifetime predictions', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-319-07176-3_2.

Balfour, B. (Hubspot) (2015) 'Growth Is Good, But Retention Is 4+Ever'. Youtube. Available at: https://www.youtube.com/watch?v=ch7aps2h8zQ.

Ballantyne, D., Christopher, M. and Payne, A. (2003) 'Relationship marketing: Looking back, looking forward', *Marketing Theory*. doi: 10.1177/1470593103003001009.

Berry, L. L. (1983) *Relationship Marketing*, *Emerging Perspectives of Services Marketing*. doi: 10.1300/J047v16n02.

Biau, G., Devroye, L. and Lugosi, G. (2008) 'Consistency of Random Forests and Other Averaging Classifiers', *J. Mach. Learn. Res.* doi: 10.1145/1390681.1442799.

Blagus, R. and Lusa, L. (2015) 'Joint use of over-and under-sampling techniques and cross-validation for the development and assessment of prediction models', *BMC Bioinformatics*. doi: 10.1186/s12859-015-0784-9.

Blattberg, R. C., Kim, B.-D. and Neslin, S. A. (2008) 'Database Marketing: Analyzing and Managing Customers', *Springer Science & Business Media*. doi: 10.1007/978-0-387-72579-6.

Bolton, R. N. (1998) 'A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction', *Marketing Science*. doi: 10.1287/mksc.17.1.45.

Bramer, M. (2013) *Principles of Data Mining*. 2nd edn. London: Springer. doi: 10.1007/978-1-4471-4884-5.

Breiman, L. (1984) 'CART: Classification and Regression Trees', *The Top Ten Algorithms in Data Mining*. doi: 10.1201/9781420089653.ch10.

Breiman, L. (2001) 'Random forests', *Machine Learning*. doi:

10.1023/A:1010933404324.

Buckinx, W. and Van Den Poel, D. (2005) ‘Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting’, *European Journal of Operational Research*. doi: 10.1016/j.ejor.2003.12.010.

Burez, J. and Van den Poel, D. (2007) ‘CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services’, *Expert Systems with Applications*. doi: 10.1016/j.eswa.2005.11.037.

Burez, J. and Van den Poel, D. (2009) ‘Handling class imbalance in customer churn prediction’, *Expert Systems with Applications*. doi: 10.1016/j.eswa.2008.05.027.

De Caigny, A., Coussement, K. and W. De Bock, K. (2018) ‘A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees’, *European Journal of Operational Research*, 269(2), pp. 760–772. Available at: https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/abs/pii/S0377221718301243.

Castro, E. G. and Tsuzuki, M. S. G. (2015) ‘Churn Prediction in Online Games Using Players’ Login Records: A Frequency Analysis Approach’, *IEEE Transactions on Computational Intelligence and AI in Games*. doi: 10.1109/TCIAIG.2015.2401979.

Cheney, C. (2016) *Understanding Churn: Why Do Customers Unsubscribe?*, *Promotional Marketing*. Available at: https://www.chiefmarketer.com/understanding-churn-why-do-customers-unsubscribe/ (Accessed: 10 September 2018).

Christian Grönroos (2008) *Service management och marknadsföring*. Malmö: Liber AB.

‘Churn’ (2018) *Gale Encyclopedia of E-Commerce*. Available at: https://www.encyclopedia.com/sports-and-everyday-life/food-and-drink/food-and-cooking/churn.

Crié, D. (2002) ‘When should a customer be defined as “lapsed”?’, *Interactive Marketing*.

Ellis, S. and Brown, M. (2017) *Hacking Growth: How Today’s Fastest Growing Companies Drive Breakout Success [Kindle Version]*. Currency. Available at: https://books.google.se/books/about/Hacking_Growth.html?id=WMl8DgAAQBAJ&redir_esc=y.

Farr, C. (2015) *Why Homejoy Failed*, *Wired*.

Frank, B. (Radford U. and Pittges, J. (Radford U. (2009) *Analyzing Customer*

Churn in the Software as a Service (SaaS) Industry. Radford University. doi: 10.15713/ins.mmj.3.

Google Developers (2018) *Classification: True vs. False and Positive vs. Negative*. Available at: https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative (Accessed: 15 November 2018).

Greis, N. P. and Gilstein, C. Z. (1991) 'Empirical Bayes methods for telecommunications forecasting', *International Journal of Forecasting*. doi: 10.1016/0169-2070(91)90053-X.

Gummesson, E. (1987) 'The new marketing-Developing long-term interactive relationships', *Long Range Planning*. doi: 10.1016/0024-6301(87)90151-8.

Gupta, S., Lehmann, D. R. and Stuart, J. A. (2004) 'Valuing Customers', *Journal of Marketing Research*. doi: 10.1509/jmkr.41.1.7.25084.

Hassouna, M. *et al.* (2015) 'Customer Churn in Mobile Markets: A Comparison of Techniques', *International Business Research*. doi: 10.5539/ibr.v8n6p224.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning. Data Mining, Inference,and Prediction., Springer Series in Statistics Trevor*. doi: 10.1177/001112877201800405.

Ho, T. K. (1995) 'Random Decision Forests Tin Kam Ho Perceptron training', in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*. doi: 10.1109/ICDAR.1995.598994.

Höst, M., Regnell, B. and Runesson, P. (2007) *Att Genomföra Examensarbete*. Lund: Studentlitteratur.

Huang, Y. *et al.* (2015) 'Telco Churn Prediction with Big Data', in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data - SIGMOD '15*. doi: 10.1145/2723372.2742794.

Hughes, A. M. (2015) *Churn reduction in the telecom industry*, *Direct Marketing News*. Available at: http://www.dbmarketing.com/telecom/churnreduction.html (Accessed: 24 August 2018).

Hung, S.-Y., Yen, D. C. and Wang, H. (2006) 'Applying data mining to telecom churn', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2005.09.080.

ICT Data and Statistics Division (2017) *ICT Facts and Figures*, *International Telecommunication Union*. doi: 10.1787/9789264202085-5-en.

Ivanonv, S. (RetainKit) (2018) *Identifying churn drivers with Random Forests*. Available at: https://blog.slavv.com/identifying-churn-drivers-with-random-forests-65bad0193e6b (Accessed: 6 September 2018).

Jackson, B. B. (1985) ‘Build Customer Relationships that Last’, *Harvard Business Review*.

Jahromi, A. T. *et al.* (2010) ‘Modeling customer churn in a non-contractual setting: The case of telecommunications service providers’, *Journal of Strategic Marketing*. doi: 10.1080/0965254X.2010.529158.

James, G. *et al.* (2013) *An Introduction to Statistical Learning*. 1st edn. Springer.

Keramati, A. *et al.* (2014) ‘Improved churn prediction in telecommunication industry using data mining techniques’, *Applied Soft Computing Journal*. doi: 10.1016/j.asoc.2014.08.041.

Khan, M. R. *et al.* (2015) ‘Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty’, in *Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015*. doi: 10.1109/BigDataCongress.2015.107.

Kim, K., Jun, C. H. and Lee, J. (2014) ‘Improved churn prediction in telecommunication industry by analyzing a large network’, *Expert Systems with Applications*. doi: 10.1016/j.eswa.2014.05.014.

Kim, M.-J., Kim, J. and Park, S.-Y. (2017) ‘Understanding IPTV churning behaviors: focus on users in South Korea’, *Asia Pacific Journal of Innovation and Entrepreneurship*, 11(2), pp. 190–213. doi: https://doi.org/10.1108/APJIE-08-2017-026.

Kirui, C. *et al.* (2013) ‘Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining’, *IJCSI International Journal of Computer Science Issues*. doi: 10.1073/pnas.94.25.13661.

Kohavi, R. (1995) ‘A study of cross-validation and bootstrap for accuracy estimation and model selection’, *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. doi: 10.1067/mod.2000.109031.

Kohavi, R. and John, G. H. (1997) ‘Wrappers for feature subset selection’, *Artificial Intelligence*. doi: 10.1016/S0004-3702(97)00043-X.

Kotsiantis, S. B., Kanellopoulos, D. and Pintelas, P. E. (2006) ‘Data preprocessing for supervised leaning’, *International Journal of Computer Science*. doi: 10.1080/02331931003692557.

Ling, C. X., Huang, J. and Zhang, H. (2003) ‘AUC: A statistically consistent and more discriminating measure than accuracy’, in *IJCAI International Joint Conference on Artificial Intelligence*. doi: 10.1039/P29880000711.

Louppe, G. *et al.* (2013) ‘Understanding variable importances in forests of randomized trees’, *Neural Information Processing Systems*. doi: NIPS2013_4928.

Madden, G., Savage, S. J. and Coble-Neal, G. (1999) ‘Subscriber churn in the Australian ISP market’, *Information Economics and Policy*. doi: 10.1016/S0167-6245(99)00015-3.

Manning, C. D. and Raghavan, P. (2009) ‘An Introduction to Information Retrieval’, in *Online*. doi: 10.1109/LPT.2009.2020494.

Milošević, M., Živić, N. and Andjelković, I. (2017) ‘Early churn prediction with personalized targeting in mobile social games’, *Expert Systems with Applications*. doi: 10.1016/j.eswa.2017.04.056.

Mitrović, S. *et al.* (2018) ‘On the operational efficiency of different feature types for telco Churn prediction’, *European Journal of Operational Research*. doi: 10.1016/j.ejor.2017.12.015.

Morwitz, V. G. and Fitzsimons, G. J. (2004) ‘The Mere-Measurement Effect: Why Does Measuring Intentions Change Actual Behavior?’, *Journal of Consumer Psychology*. doi: 10.1207/s15327663jcp1401&2_8.

Mutanen, T. (2006) *Customer churn analysis–a case study*, *Technical Research Centre of Finland (VTT)*.

Óskarsdóttir, M. *et al.* (2017) ‘Social network analytics for churn prediction in telco: Model building, evaluation and network architecture’, *Expert Systems with Applications*. doi: 10.1016/j.eswa.2017.05.028.

Periáñez, Á. *et al.* (2016) *Churn Prediction in Mobile Social Games: Towards a Complete Assessment Using Survival Ensembles*. doi: 10.15713/ins.mmj.3.

Portela, S. P. and Menezes, R. (2010) ‘An Empirical Investigation of the Factors that Influence the Customer Churn in the Portuguese Fixed Telecommunications Industry: A Survival Analysis Application’, *The Business Review, Cambridge*.

Powers, D. M. W. (2011) ‘Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation’, *Journal of Machine Learning Technologies*. doi: 10.1.1.214.9232.

Qin, Z. and Tang, Y. (2013) *Uncertainty Modeling for Data Mining*. eBook. Hangzhou: Springer.

Reichheld, F. F. (1996) ‘The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value.’, *Academy of Management Perspectives*. doi: 10.5465/AME.1996.9603293227.

Reichheld, F. F. and Sasser, W. E. (1990) ‘Zero defections: quality comes to services.’, *Harvard business review*. doi: 10.1016/j.colsurfa.2006.11.029.

Reinartz, W. J. and Kumar, V. (2000) ‘On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and

Implications for Marketing', *Journal of Marketing*. doi: 10.1509/jmkg.64.4.17.18077.

Reinartz, W., Thomas, J. S. and Kumar, V. (2005) 'Balancing Acquisition and Retention Resources to Maximize Customer Profitability', *Journal of Marketing*. doi: 10.1509/jmkg.69.1.63.55511.

Robehmed, N. (2014) 'The Year's Hottest Startups.', *Forbes*.

Rokach, L. and Maimon, O. (2005) 'Top-down induction of decision trees classifiers - A survey', *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*. doi: 10.1109/TSMCC.2004.843247.

Sarkar, D., Bali, R. and Sharma, T. (2018) 'Building, Tuning, and Deploying Models', in *Practical Machine Learning with Python*. Berkeley, CA: Apress, pp. 255–304. doi: 10.1007/978-1-4842-3207-1_5.

SAS Institute (2016) *Data Mining from A-Z: How to Discover Insights and Drive Better Opportunities*. Available at: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-mining-from-a-z-104937.pdf.

Scikit-learn (2018) *Random Forest Classifier*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (Accessed: 17 November 2018).

Shin, H. W. and Sohn, S. Y. (2004) 'Multi-attribute scoring method for mobile telecommunication subscribers', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2003.09.013.

Sifa, R. *et al.* (2015) 'Predicting Purchase Decisions in Mobile Free-to-Play Games', in *Proceedings, The Eleventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-15)*.

Statista (2018a) *Average monthly churn rate for wireless carriers in the United States from 1st quarter 2013 to 1st quarter 2018*. Available at: https://www.statista.com/statistics/283511/average-monthly-churn-rate-top-wireless-carriers-us/ (Accessed: 3 September 2018).

Statista (2018b) *Average thee month user retention and churn rate of mobile apps worldwide as of 2nd half 2017*. Available at: https://www-statista-com.ludwig.lub.lu.se/statistics/384224/monthly-app-launches-churn/ (Accessed: 3 September 2018).

Subramanya, K. B. and Somani, A. (2017) 'Enhanced feature mining and classifier models to predict customer churn for an E-retailer', in *Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering*. doi: 10.1109/CONFLUENCE.2017.7943208.

Tape, T. G. (University of N. (2018) *The Area Under an ROC Curve*. Available at: http://gim.unmc.edu/dxtests/roc3.htm (Accessed: 16 November 2018).

Telegeography (2017) *Telegeography Report*.

UC Allabolag AB (2018) *Rebtel Networks AB Bokslut*. Available at: https://www.allabolag.se/5566803622/bokslut (Accessed: 28 August 2018).

Vafeiadis, T. *et al.* (2015) 'A comparison of machine learning techniques for customer churn prediction', *Simulation Modelling Practice and Theory*. doi: 10.1016/j.simpat.2015.03.003.

Valverde-Albacete, F. J. and Peláez-Moreno, C. (2014) '100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox', *PLoS ONE*. doi: 10.1371/journal.pone.0084217.

Verbeke, W. *et al.* (2011) 'Building comprehensible customer churn prediction models with advanced rule induction techniques', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2010.08.023.

Wei, C. P. and Chiu, I. T. (2002) 'Turning telecommunications call details to churn prediction: a data mining approach', *Expert Systems with Applications*. doi: 10.1016/S0957-4174(02)00030-1.

Yu, X. *et al.* (2011) 'An extended support vector machine forecasting framework for customer churn in e-commerce', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2010.07.049.

Zhang, X. *et al.* (2012) 'Predicting customer churn through interpersonal influence', *Knowledge-Based Systems*. doi: 10.1016/j.knosys.2011.12.005.

Zhu, B., Baesens, B. and vanden Broucke, S. K. L. M. (2017) 'An empirical comparison of techniques for the class imbalance problem in churn prediction', *Information Sciences*. doi: 10.1016/j.ins.2017.04.015.

1 Appendix

1.1 Dataset description

Table 6: Description of model variables

Feature	*mean*	*std*	*min*	*25%*	*50%*	*75%*	*max*
TotalDuration	738,8	815,97	0,02	228,1	505,1	967,7	23411,72
DurationTrend30	1,50	221,24	1,00	0,35	0,05	0,32	67816,71
CountCalls	67,80	66,22	1,00	24,00	49,00	90,00	1288,00
CallTrend30	0,21	2,16	1,00	0,31	0,04	0,29	176,00

NumberOfBNumbers	9,86	7,00	1,00	5,00	8,00	13,00	98,00
BNumbersTrend30	0,13	0,84	0,98	0,29	0,00	0,33	48,00
CallTrend15	3,33	37,27	-1,00	0,33	0,01	0,38	1819,00
DurationTrend15	4,24	51,61	-1,00	0,38	0,04	0,42	4502,55
DaysCallsMade	18,96	8,64	1,00	12,00	20,00	27,00	30,00
NbrShortCalls	7,07	9,28	0,00	2,00	4,00	9,00	513,00
PropShortCalls	0,10	0,09	0,00	0,05	0,09	0,13	1,00
FailedCalls	33,08	46,16	0,00	8,00	20,00	41,00	3273,00
PropFailedCalls	0,62	1,52	0,00	0,23	0,40	0,67	176,38
MonthsSinceSignup	-23,25	20,60	-212	-30,0	-19,00	-10,00	-2,00
SumTotBilling	9,91	4,20	0,00	10,00	10,00	10,00	460,00
NbrPayments	1,00	0,41	0,00	1,00	1,00	1,00	46,00
DaysSinceLatestPayment	-17,31	11,01	-60,0	-25,0	-17,00	-8,00	-1,00
DaysSinceLastCall	-3,31	5,35	-30,00	-3,00	-1,00	-1,00	-1,00
DaysDiffBetweenCalls	-1,00	2,83	-59,00	-1,00	0,00	0,00	0,00
NbrOfContacts	268,73	1719,10	1,00	91,00	180,00	315,00	505559,00
Churn	0,11	0,31	0,00	0,00	0,00	0,00	1,00