Aalto University

MS-E2177
Seminar on Case Studies in Operations Research

Lifetime value modeling for strategy development

Final Report

Otto Saikkonen

Oscar Björklund

Aki Malinen

Iiro Sallinen

Juho Heimonen

May 31, 2020

Contents

1 Introduction

1.1 Background

This case study is conducted for Nordea Bank, a full-service universal bank and one of the biggest banks in Europe, with a total operating income of EUR 8.6 billion and total assets of EUR 554.8 billion in 2019 [1]. Nordea serves customers in 19 countries, including the home markets Denmark, Finland, Norway and Sweden. Nordea serves approximately 9.3 million household customers, with 3.1 million in Finland [1]. Nordea offers services in personal banking, business banking, large corporates and institutions and asset & wealth management. Table 1 shows the key figures of Nordea Group.

	Personal Banking	**Business Banking**	**Large Corporates & Institutions**	**Asset & Wealth Management**
# Customers	9 000 000	4500	1800	94 000
# Employees	7500	4500	1800	2700

Table 1: Key figures of Nordea group. Personal banking is the largest business segment of Nordea in terms of customers.

This study focuses on the insurance division of Nordea, Nordea Life & Pensions, which aims to be the leading provider of life and pension products in the Nordic countries. The subsidiary is tightly integrated with the rest of the Nordea Group, to ensure seamless offers to Nordea's customers in all key life events [3]. Nordea MyLife personal risk insurance is the most important product of Nordea life. It includes coverage for various life events such as unemployment, death, serious illness and the claim payments are predefined single lump sums or in the case of temporal unemployment, predefined monthly payments.

1.2 Customer Lifetime Value in the context of insurance industry

For the insurance industry, the premiums of customers are the profit sources while the settlement of claims and customer acquisition costs are the major expenditures. According to some literature, the insurance industry has the highest customer acquisition costs of any industry. It can cost seven to nine times more for an insurance agency to attract a new customer than to retain one[2]. The premium margin, customer relation length and the claim risk are thus the most important aspects for insurance companies for evaluating the profit of customers [4]. While investment of assets is a major source of income for insurance companies [3], we focus on the B2C interaction layer where the return of invested capital is not relevant.

Due to the high customer acquisition costs of the insurance industry, it is not meaningful to target every potential customer of an insurance product equally.

Therefore, it is important to focus customer acquisition resources and efforts towards the customers with the highest potential value in order to drive future profits. This implies creation of a metric for differentiating the value of different customers. Such metric can not only be used for customer selection and marketing resource allocation, but it can be used to analyze customer shifts between segments and to analyze the factors that may influence customer behavior. **Such metric is referred here as the Customer Lifetime Value (CLV).**

1.3 Objectives of the case study

The current main metric for measuring the value of a customer at Nordea across the different branches is the **total value of customer assets under management**. While this gives an estimate of the wealth and purchasing power, it offers a poor valuation in the context of risk insurance. The purpose of this case study is to study the concept of Customer Lifetime value (CLV) and propose better metrics for CLV in the context of risk insurance business.

Particularly, the main pre-defined objectives of this case study are:

1. Define an example measurement of CLV relevant to the business context
2. Build a predictive model for CLV
3. **Propose decision-making strategies based on the findings**

We start by taking a look at the literature and outlining the methods for evaluating customer value in different business contexts. Main outcomes of the literature review are presented in the Section 2.
We then attempt to take one possible method of predicting CLV based on the literature and make a practical example of it. This implies getting a sample data from the customer company. The process of acquiring and managing the data are described in the Section 3 while the methods and results are described in Section 4 and Section 5. Finally, we take what we learned and try to propose strategy changes based on the findings. It is important to note that in order to answer to the objective 3, simply finding a model that makes a prediction of CLV based on the data we receive, does not necessarily serve much value. Firstly, simply due to legal and data protection reasons we are not able to access all of the data that is available in Nordea. Even if such information was theoretically available under NDA contracts, releasing such data would pose unacceptable corporate level risks. Furthermore, the current way Nordea gathers data might not be sufficient for meaningful predictive CLV modeling. In order to have meaningful input for **strategy**, or *a plan of action designed to achieve a long-term or overall aim*, we need to evaluate the different phases of the project in the business context and focus on questions such as:

- What kind of CLV model is applicable to the context if all information was available?
- How does one take maximal advantage of predictive CLV modeling in

everyday business functions of a risk insurance company?

- Does the current organization and data management and collection measures allow strategy based on predictive modeling?
- What kind of systematic changes are required for implementing more robust predictive analytics?

2 Literature Review

2.1 Modeling Customer Lifetime Value

There exists no unique mathematical definition of the Customer Lifetime Value (CLV). To tailor the CLV metric for particular needs, researchers have used different approaches to the modeling and estimation of CLV. Without losing generality, one could define it as any quantitative method that is used to attach a numerical value to a customer or a customer segment that describes the value of a given customer in comparison to others. In literature, CLV is often similar to the discounted cash flow approach widely used in finance. However, there are two differences. First, CLV is typically defined and estimated at an individual customer or segment level. This allows differentiation between customers who are more profitable than others rather than simply examining average profitability [5]. Second, CLV incorporates the probability that a customer stops being a customer, or *churns*.

Further, the time horizon varies depending on the context: Some researchers have used an arbitrary time horizon or expected customer lifetime (for example [16], [17]), whereas others have used an infinite time horizon (for example [18], [11]). Based on the literature review, the most common form of CLV is defined as the present value of all future profits obtained from a customer over his or her entire lifetime with the firm [5], [6]. This is sometimes known as customer lifelong profitability. Some simpler approaches focus on net present value of customers transactions over the lifetime of a customer[7], [8]. The key difference here is that the former aims to describe the real profitability of a customer, while the latter focuses on turnover: transactions do not account the costs of retaining or acquiring the customer, nor does it consider the net margin associated to the products associated to the customer.

Gupta and Lehmon [9] propose the following basic model:

$$CLV = \sum_{t=0}^{T} \frac{(p_t - c_t)r_t}{(1+i)^t} - AC, \tag{1}$$

where
p_t = price paid by a consumer at time t,
c_t = direct cost of servicing the customer at time t,
i = discount rate or cost of capital for the firm,
r_t = probability of customer repeat buying or being "alive" at time t,
AC = acquisition cost, and
T = time horizon for estimating CLV

Blattberg and Deighton [15] suggested that growing a business can be framed as a matter of getting customers and keeping them so as to grow the value of the customer base to its fullest potential. Customers go through a lifecycle process with a firm. A firm, by investing resources, first acquires customers. Then it expects to make reasonable profits by offering products and services to these customers. Over a period of time, the firm also experiences customer attrition. The process of growing the value of the customer base to the fullest can be represented as the strategic impact of managing the CLV process over the span of a customer's lifetime with the firm (see Fig. 1 by Kumar et.al.).

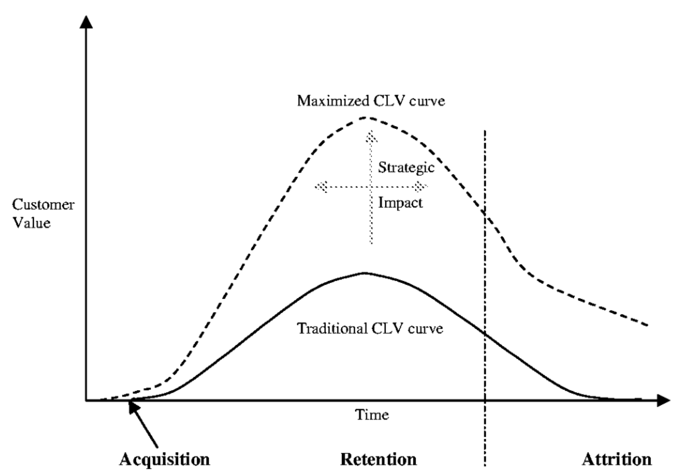

Figure 1: Managing the CLV process by strategic impact in the course of a customer's lifetime with the firm [6].

The key argument to be taken from Blattberg and Deighton is that simply considering the cash flow models based on available customer information do not consider the total value of a customer. Consequently, the parameters associated to these models should perhaps not be the only indicators to consider in business decisions. Each customer also has their own unique strategic value. Furthermore, there is no one unique way to measure the strategic value: For example, having a large non-profitable customer base in personal banking can offer significant strategic value for other business operations, because the customer relation not only offers a simple direct way to contact the customer for new sales (such as insurance), it also serves as a customer data point for big data applications. The customer satisfaction can be seen as source of strategic value, since very satisfied customers have been shown to drive sales without active involvement of the company. Such strategic value could also be estimated using key performance indicators such as Net Promoter Score [26].

Prior to considering what kind of model to choose for modelling CLV, it needs to be understood how the CLV will be used. Hwang [10] propose marketing strategies according to segmentation based on the customer value. After determining the CLV of individual customers, it is possible to segment the customers into groups based on the CLV, churn or any other aspect of the customer, or combination of aspects. There are various algorithms such as PCA and K-Means Clustering that can be used for segmentation of customers. The goal of the segmentation is to divide the customers into groups of distinct characteristics, to be used when developing marketing strategies for each group. The goal of this process, resulting in marketing strategies, is to identify which group of customers, based on their characteristics, is or will be most profitable for the business and where to target marketing resources.

However, clustering that combines a collection of different data points is often some kind of a compromise between resources available for marketing and number of clusters. When the amount of data and analytical methods allow us to get individual level data, does it still make sense to use this kind of micro-segmentation if we can target individuals directly now that technology has enabled us to do so? Most likely there is a purpose for micro segments in the future also, however we argue that the number of segments will increase and they are increasingly overlapping. For an example, while the group of "home owners" merges together a segment of customers who could me more prone to reacting to marketing

Hwang uses a conceptual framework, which evaluates customer value from three viewpoints - current value, potential value and customer loyalty. Fig. 2 shows the conceptual framework of Hwang [10].

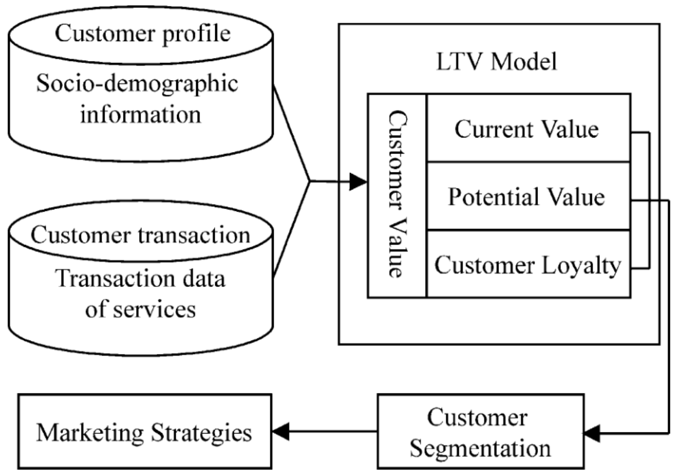

Figure 2: Conceptual framwork for modeling and using CLV to implement marketing strategies [10].

2.1.1 Current value

Current value is defined as a profit contributed by a customer during a given period [10] from present to the future, not as a cumulative value from the past to the present. The current value can be calculated based on the current products and services the customer is engaged with at the moment.

2.1.2 Potential value

Potential value is important to consider when calculating customer value due to cross-selling and up-selling. Potential value of customers can be defined as expected profits that can be obtained from a given customer when a customer uses the services [13]. The equation to evaluate potential values is

$$\text{Potential value}_i = \sum_{j=1}^{n} \text{prob}_{ij} \times \text{profit}_{ij} \tag{2}$$

Prob_{ij} is the probability that customer i will use the service j among n-optional services. Profit_{ij} means the profit that a company can receive from the customer i who uses the optional service j. In other words, the equation above means expected profits from a particular customer who uses optional services provided by the company. Profit_{ij} means the expected value when a company provides a customer with a certain optional service [13].

The probabilities associated with sales of additional products can be hard to estimate and requires careful analysis of the success of different sales channels. However, the potential value analysis can also be performed disregarding the acquisition probabilities by applying predictive CLV modeling to answer conditional questions such as: Based on the existing data, what is the lifetime net monetary value of the customer if he/she buys product x?

2.1.3 Customer churn prediction

Churn is a marketing-related term characterizing a consumer who leaves a service relationship with a company. If the company wants to prevent the consumer from leaving, a retention action is required. The population of interest is the customers that have already been acquired.

The condition under which a customer has to be considered as decreasing his/her loyalty, and hence churning, can be problematic to define, as outlined in [12]. The issue in a competitive environment is that most people have more than one supplier. For instance, in retail banking, a customer could have a current account in a first bank and a mortgage loan in another. Most people have several current accounts even if they do not use them.

Kim [13] defines customer loyalty as

$$\text{Customer Loyalty} = 1 - \text{Churn rate}. \tag{3}$$

Here, churn describes the number or percentage of regular customers who abandon a relationship with a service provider. Customer loyalty can be a measure of customer retention. Kim [13] measure the leaving probability for each customer to calculate the churn rate, using data mining techniques. This study takes a similar approach.

The customer churn is therefore closely related to to the customer retention rate and loyalty. Hwang [10] defines the customer defection the hottest issue in the highly competitive wireless telecom industry. Their Lifetime Value (LTV) model suggests that churn rate of a customer has strong impact on the LTV because it affects the length of service and therefore future revenue. A deeper study on customer churn prediction in the retail banking industry was done in [14].

2.2 Development of the CLV metric for the context of risk insurance

We now focus on the monetary CLV based on cash flow analysis. According to Hwang [10], The CLV of an individual customer is mainly dependent on two factors: the customer profile (for example age, gender and general health) and customer behaviour. Based on information about the individual customer, it is possible to calculate the CLV for every individual customer separately. This idea can be extended as follows: Giving the company can acquire enough data to estimate the key factors driving profitability, such as retention rate and monthly purchases, the company can calculate the potential CLV for a given product.

Furthermore, if the company can develop enough data to estimate the costs of acquisition through each sales channel for each customer profile, the company can create a complete prediction of the potential CLV of a potential customer.

The model in equation (1) fails to address the complete value transaction chain of an risk insurance business-customer relation. It is designed to the context of products and services where the "direct cost of servicing" are easily defined, for an example as a sum of fixed and variable costs associated to the products in question. In the insurance industry, the whole core business revolves around managing and predicting the risk associated to insurance claims. Thus, we argue that a complete description of the monetary CLV of a customer in risk insurance context must include the best available information of the risks associated to the customer. Moreover, insurance contracts often include covers for various kind of events, all of which have different associated risks for a given customer. To address these points, we propose to extend the equation (1)

$$CLV = \sum_{t=0}^{T} (\frac{(p_t - c_t)}{(1+i)^t} r_t - \sum_{k=0}^{K} cl_k * P(cl_{kt})) - AC, \tag{4}$$

where $\sum_{k=0}^{K} cl_k * P(cl_k, t)$ describes the expected value of claim payments during time period t, cl_k is the claim payment associated to event k and $P(cl_k, t)$ describes the expected risk of event k during time period t. The equation works in the context of Nordea MyLife, because the insurance only covers a finite discrete set of events and the claim payments are lump sums that are predefined for a given event based on the age, claim sum and information about smoking.

If we talk about the current value of an existing customer, the acquisition cost has been realized already and it should not be considered. However, if we talk about the *potential customer value*, the acquisition cost is a significant factor. Extending the concept of individually determined risk and retention functions, also the probability of lead to sales varies from person to person and thus, giving that there is sufficient information about the potential customers available, the probability of a sale (or the conversion rate) can also be predicted for different customer groups for a given sales channel. Thus, theoretically if the cost of sales attempt is known, it is possible to predict the potential net monetary value for a given customer and a product.

In order to choose an appropriate model for the case study, we have to consider what kind of data is available and what kind of methods are meaningful in the context. In this case, the corporate structure of Nordea combining both personal banking and personal insurance offers an interesting opportunity to study the joint customer base. This argument is motivated by the following underlying assumptions:

- By creating an intersection of the existing customers of the banking and insurance branches, we can increase the available attributes for advanced analytics.

- The information related to personal finances can be used to analyze the purchase power of a customer, which by assumption correlates with the value of a potential insurance contract monthly premium.
- Theoretically, the management of personal finances can be used to observe risk averse/risk seeking behaviour which correlates with events related to risk insurance claims
- We assume that the average retention rate of customers or a customer segment is affected by various hidden or visible attributes related to personal finances
- By analysing the time series related to personal finance, one can identify potential life events of a customer which are associated to sudden churn: If for an example the momentary purchase power of a customer is suddenly reduced due to momentary circumstances, the company can first make an analysis of the long term current and potential lifetime value. This information is then used in deciding whether to passively let the customer churn or actively try to prevent churn by offering flexibility in payments.

• By getting a description of the data on corporate level and analysing the process of acquiring data for CLV modeling, we can analyse the applicability of the current data management system for advanced data driven decision making strategy.

In the case of our case study, we do not have enough data or time to formulate the risk function. Moreover, we do not have enough information regarding the costs associated to acquiring and maintaining the customers. Thus we propose the use of a reduced version of the (4) where we only study the revenue of a customer based on the retention rate and the monthly premium.

$$CLV = \sum_{t=0}^{T} \frac{p_t}{(1+i)^t} r_t \tag{5}$$

Thus, in order to predict the CLV of a customer with this equation, we only need to consider the estimated monthly premium and the estimated retention rate of a customer.

3 Data

We note that the data structures, attributes or other reflections of the data management used in this case study do not represent the real underlying data management at Nordea Group. All of the references to data structures and conclusions made here are based on the simulated "case study" data.

In order to explore different CLV models and the process of using data based CLV modeling, we requested a description of the customer data that is managed in the Nordea corporation. Particularly, we requested to get sample data of different branches of Nordea. Based on the description we received, we made a data query request to get a sample combining the insurance customers and the banking customers.

The data we received was customer data and product data. The customer data is from four different internal sources from Nordea. The product data came from just one source.

We are not consuming all of Nordeas data but rather a fraction of it. At the start of the collaboration we agreed on picking attributes if interest from a data catalog. This allowed for us to push for getting what we really needed and Nordea to counter with their version because of security reasons. In the end we would get close to what we asked for in terms of attributes.

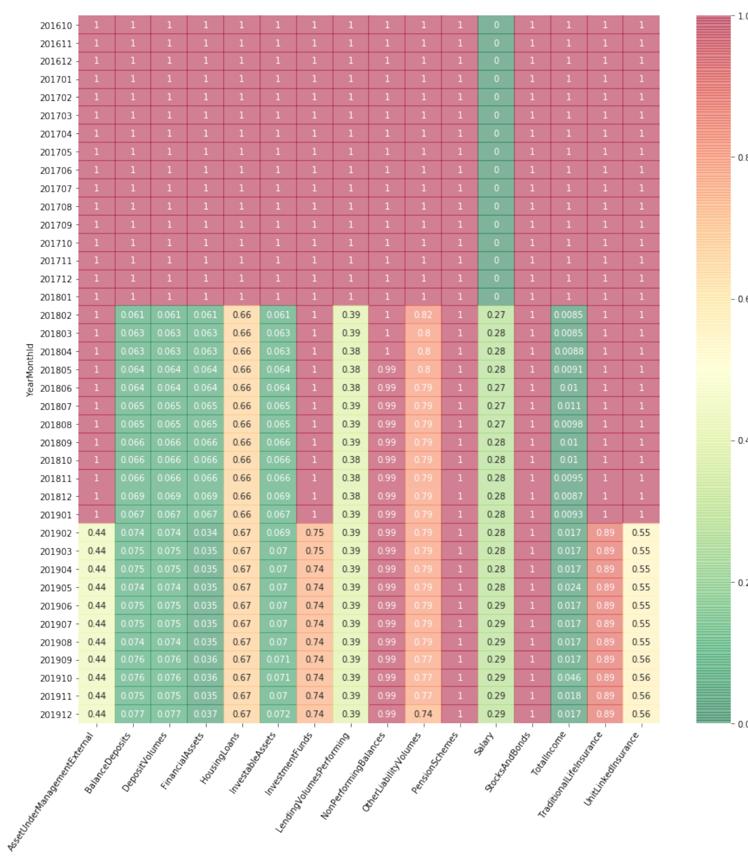

Figure 3: Population of attributes in source 1

The data in general has many disconcerting data quality issues. Some of these are visualized in Figure 3. The plot shows how much of a given attributes fields are zero for all customers. We can see that most of the data is just zeros. This makes the analysis difficult because the lengths of the time series are impacted heavily and we can not make distinctions between actual zeros and non-populated fields i.e. every zero has to either be handled as a zero or as non-populated or null. In addition to the fundamental validity of and consistency of the data, we identified other weaknesses in the way the data has been collected: For an example, the data source collecting information of contracts does not clearly point of situations where the contract was upgraded to another, possibly more valuable contract. Without considering information about this kinds of events, uninformed analyst interprets this as two different short contracts and this greatly affects the estimated churn rate.

While simulating a data set is not inherently inapplicable, it seems the data we received was not simulated properly for our purposes. For example customers age in different points in time is not even close to linear. This of course should be exactly one year in a year, but that is not case with our data. This effect is present in almost all attributes that should have a similar perfect correlation e.g. lengths of contracts or total income and salary. It should be noted that this may not be only because of the simulation of the data. Other preexisting data quality issues may be the root cause but that is something we can not determine. We also assume that the addition of variance did not consider the correlations between the separate data sources.

Another problem that might affect the performance of analysis is the documentation of data: In our case, the data came without any documentation and the attributes were not always unambiguous. This is especially challenging in the case of abbreviated status attributes that are common in databases. The more of data is left for interpretation, the more prone the analysis is to human errors. This is not just a problem for external consulting or case studies, but in general: Even if the attributes are extremely clearly named, even the best analysts need to know exactly how the data was constructed in order to make conscious analytics.

Due to the considerable problems with the quality and consistency of the data together with the fact that the data set simulated in order to avoid identifiability, the results presented in **Section 5** should be taken with sufficient reservations.

4 Methods

In this chapter, we introduce tools that can be used in the process of estimating CLV or variables needed to calculate the CLV. Particularly, we present The Kaplan Meier estimator that was the main method that we use to model the customer loyalty (or churn rate) and present the conceptual basis of the two predictive models we tried to train to predict the customer CLV. We also present the classification model that can be used to classify between Mylife customers and customers without Mylife product. The goal of this classification task is to spot feature that characterize Mylife customers.

4.1 Kaplan Meier estimate

The Kaplan Meier estimate is one of the most important tools when estimating the survival function [20] that tells the probability of a customer still being "alive" as a function of time. The estimator is used for instance in medical research to estimate patient lifetime and in the estimation of service times of spare parts. In the context of this project the survival function is used for analysis of the customer loyalty. The Kaplan Meier estimate uses contract lifetime data and the estimation of the survival function is given by

$$\hat{S}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}), \tag{6}$$

where i stands for an index for discrete time points t_i, d_i is the number of "deaths" at time t_i, and n_i is the number of individuals that have known to survived to time t_i. One can notice that the function given by the estimator is decreasing until the time that is the length of the longest ended lifetime. After this point in time, the function is constant. This limits the length of the time interval we can hope to get reliable results: if the data does not include ended contracts with length longer than 10 years, we can not estimate the survival function for over 10 years.

In this project, we use an open source implementation for estimating the Kaplan Meier curve. The python library *lifelines* is used for this purpose [19].

4.2 Gradient boosting method

Gradient boosting is machine learning technique, which uses weak learners (e.g., decision trees) to emsemble a prediction model for regression or classification problems. Weak learners are built in a stage-wise way, and it generalizes other boosting methods by allowing optimization of an arbitrary loss function. In the algorithm, *pseudo-residuals* r_{im}

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial L(x_i)}\right], \tag{7}$$

where $L(\cdot)$ is a differentiable loss function [22]. Then a multiplier γ_m is solved from

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma_m h_m(x_i 1), \tag{8}$$

where h_m is the fitted weak learner with $r_i m$. Finally, the model is updated as

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x). \tag{9}$$

Gradient boosting models powerful are that there is a lot of flexibility in tuning the hyperparameters, the data needs little pre-processing (can handle categorical and numerical values as is) and it handles missing values, and thus the data does not need any imputation.

4.3 Artificial neural networks

Artificial neural networks (also known as deep learning models) are powerful tools for constructing and training complex models with large data sets. The basic idea of artificial neural networks is to mimic a mammal brain that consist of billions of neurons. The neural networks can be utilized in any machine learning tasks such as classification, regression and clustering and they are especially useful in fields such as computer vision and language processing. [21]

A simple neural network consists of many fully connected layers. Every layer consists of neurons that are connected to all neurons of the previous layer and the next layer. A single neuron passes forward the weighted sum of its inputs though a non-linear activation function. An important innovation in the training of the neural networks is the so called backpropagation. It enables the automatic gradient calculation of complex models so that the training can be carried out using gradient decent type methods.

In this project, we seek to apply a neural network model to predict the customer CLV from different kinds of customer features. This is done using the open source Python library *Pytorch* that provides easy to use tools for constructing and training custom neural networks. The library also enables the use of graphics processing unit (GPU) in the training process, which leads to significant improvements in training speed.

4.4 Predictive model for CLV

The reasoning behind predicting customer CLV is simple: if a model learns to predict customers' CLVs, we can look at the interactions the model has learned to get insights on what kind of customers have high CLV. This methodology comes with limitations, since 1) we have to ensure that the model performs well enough and 2) even if the model performs well, we cannot explicitly claim that our analysis holds in the real world - it is just what the model has learned. However, with a good amount of high quality data, we assume that 2) our findings line up with the real world.

In Sections 4.2 and 4.3, we introduced two predictive models, *gradient boosting tree* and *deep neural netword*. We will use these two models to try to predict MyLife customers' CLV and use the model which performs better.

4.5 Classification model for detecting customers

Predicting CLV is a regression problem which is quite problematic if data is limited. In order to get something out of our analysis, we try to reformulate our regressions problem to a classification problem. Compared to a regression problem, where we try to predict a continuous value, classification problem is more simpler: try to predict in which of predetermined classes does the given data point belong to. Moreover, we formulate our problem as a binary classification problem, i.e., the model has to indicate if given data point belongs to a single class or not. The single class we use are MyLife customers, and therefore, we try to predict the probability that a given customer has MyLife product.

The model we use in this problem is a gradient boosting classifier, due to its performance and quick-to-learn abilities.

4.6 Feature analysis

For the feature analysis, we will use the `SHAP` package available from PyPI. *SHAP* is abbreviation from *SHapley Additive exPlanations*, which is a game theoretic approach to explain the output of any machine learning model, and it connects optimal credit allocation with local explanations using the classic Shapley calues from game theory [25].

With the `SHAP` package's explainer module, we are able to get a *shap value*, for each feature for every prediction. Moreover, for a single data point,

$$\exp\{\mu\} + \exp\{\sum_i s_i\} = \exp\{p\}, \tag{10}$$

where μ is baseline prediction, i.e., the average of people who have MyLife, s_i is the shap value for feature i and p is the prediction made for the data point. In its essence, shap values will tell us if a feature value had positive or negative impact on a prediction, as well as the impacts magnitude. This allows us to analyze feature interactions in detail.

5 Results

The goal of this project was to find a suitable metric for CLV in the context of insurance customers that could be then predicted from other customer features using machine learning. However, the models we tried to implement did perform sufficiently. Neither of our approaches, the neural networks or gradient boosted tree could learn sufficiently from the data. This is likely due to the fact that the data we received was simulated and had too much noise. This issue is discussed in more detail in chapter 6.2.

The objective we landed on after this setback was to predict a given customers probability of having a certain product. This allows for determining how similar a customer is compared to other products' customers. Then based on the probalilities we analyze the trends and features that are present with a products' customer base.

5.1 Suggestion for a CLV in the context of life insurance customers

Based on the literature reviewed in section 2, we suggest a measure for customer lifetime value consisting of three main parts: the turnover from the customer, the costs related to the customer and the expected loyalty of the customer. Let us next dive in detail into these one by one. Furthermore, the CLV proposed here is suitable for customers that already have life insurance products. The prediction of potential value for potential customer is discussed in the next chapters.

5.1.1 Turnover from customer

In the context of life insurance business, the turnover produced by a single customer can be considered to be relatively deterministic: it consists of the monthly or annual premiums that the customer pays according to the contract. The contracts are usually quite long and the price is stable. The size of the premium is for large part determined based on the size of the cover amount and age of the customer. Here, the cover amount refers to the monetary compensation received in case of death or illness. In addition, things such as whether the persons smokes or not, can affect the price. However, insurance companies are not allowed to discriminate.

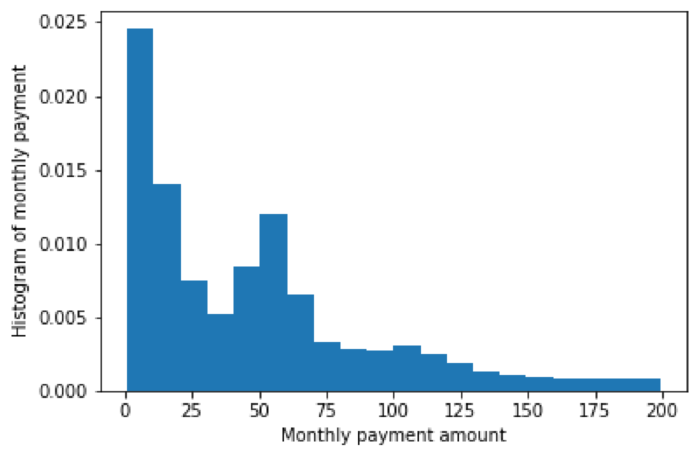

Figure 4: Histogram of the monthly insurance premiums. The values are simulated.

As the premiums are almost deterministic, their cumulative amount can be easily calculated when considering the customer lifetime value. Even though CLV stands for the value over the customer lifetime value, we recommend that the premiums should not be considered for many decades forward in time. Although some customerships might last for that time, we consider that the near future should have more weight in the customer value as it can be more easily predicted. Furthermore, discounting with some reasonable internal rate of return should be done when summing up the future premiums. Histogram of the used monthly premiums is presented in figure 4.

5.1.2 Expected costs due to the customer

The costs of a single life insurance customer can be divided into two parts: costs for selling the insurance product to the customer(marketing, contacting customers, etc.) and cost that occurs if the person dies and the cover amount have to be paid. In the context of this project, we neglect the first one, as it can be difficult to pinpoint certain marketing costs to certain customers. Moreover, the data we used did not cover these costs. Hence, the risk of death (and cover amount being realized) should be the main source of costs when considering the customer value.

The risk of death is affected by many things, the most significant being age, gender and general health. We suggest that open source death statistics should be used when considering the expected costs of realized compensations. Example of such death probabilities are presented in figure 5. One can notice that on average, males have higher probability of death than females and that the risk of death starts to increase after the age of 40 and accelerates after the age of 70 years.

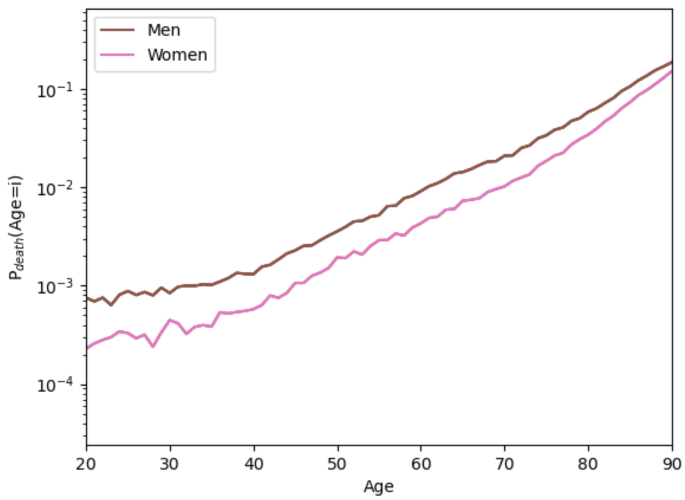

Figure 5: The probability to die during one year given persons age. The mortality data is obtained as five year averages from Statistics Finland open source data [24].

5.1.3 Customer loyalty

While the expected incomes and the cover amount in case of death are fixed, the customer loyalty plays a significant role in the CLV. In order to analyze the customer loyalty in different segments, we use the data of the contracts to estimate the customer survival function. Survival function tells the probability that an individual is still a customer as a function of time. The function has the value one at time zero, when the customership starts and the function is decreasing by nature.

As discussed in the chapter 4.1, we use the Kaplan Meier estimate to form the survival functions. Some example survival functions are shown in the figure 6. The figure shows longer customerships in younger customer segments. With the customers older than 75, the decreasing survival function is due to limited human life. Furthermore, women seem to be more loyal customers than men.

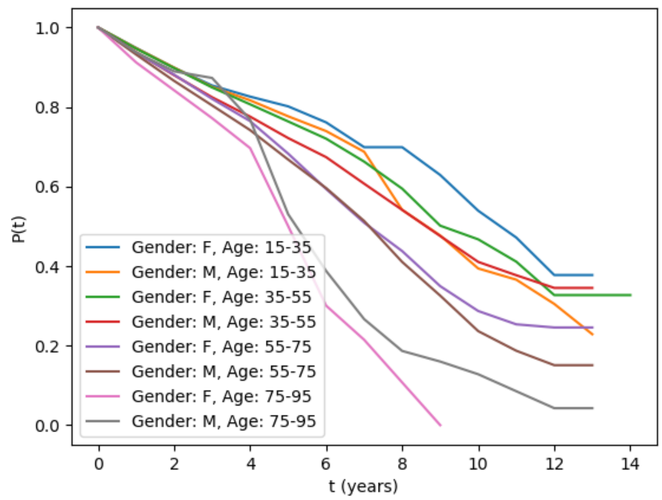

Figure 6: Survival functions for different customer segments.

5.1.4 Total CLV

We suggest a CLV measure that takes account all the three factors described in the previous chapters. This is done by calculating the expected premiums from the customers using death probability and survival function and by calculating the expectation value of the costs due to the cover amounts. After the customer segment specific probability values are known, the calculation becomes a straightforward calculation of expectation value.

In practice, this was done so that customers were classified into segments based on gender and age. For each group, a separate survival function was estimated using the Kaplan Meier estimate. Notice that some other segmentation of customers could also be possible. The CLV was calculated based on only the monthly premiums and the survival function. This prediction was done for next five years. The possible costs due to the customer were neglected because the data we used did not contain the cover amounts data. This leads to having a CVL measure that has large weight on the monthly premiums. This is not necessary a bad thing if we assume that the profit is a constant proportion of the turnover. The CLV values calculated for the Nordea life insurance customers are shown in the figure 7. The CLV value that we used can be characterized by the formula

$$\mathrm{CLV} = \sum_{i=0}^{4} \frac{p_i A_i}{(1+r)^i}, \tag{11}$$

where p_i is the probability that the person is still a customer at year i, A_i is the annual premium and r is the interest rate used for discounting.

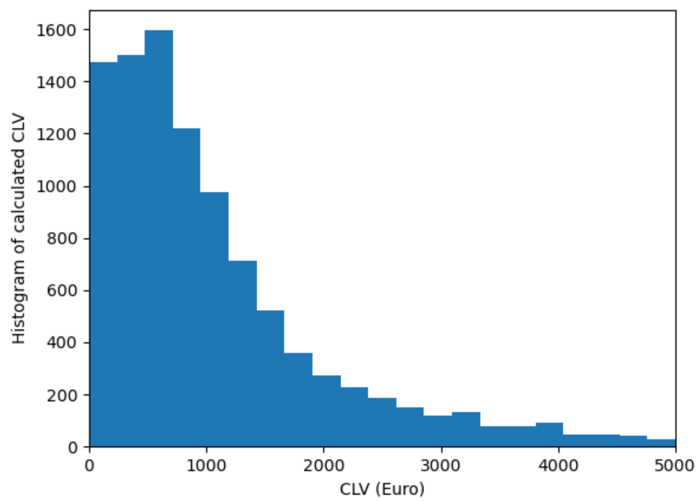

Figure 7: Histogram of the calculated CLV values for Nordea life insurance customers.

5.2 Results from classifying MyLife customers

Our model was able to correctly classify MyLife customers well. This can be seen from Figure 8, where we can see that the density distributions of predicted probabilities for MyLife and non-MyLife customers. In addition, the ROC curve in Figure 9 looks good. The area under ROC curve was 0.95, which is excellent. The precision of the test set was 0.78 and recall 0.28, which strenghtens our argument that the model performs well.

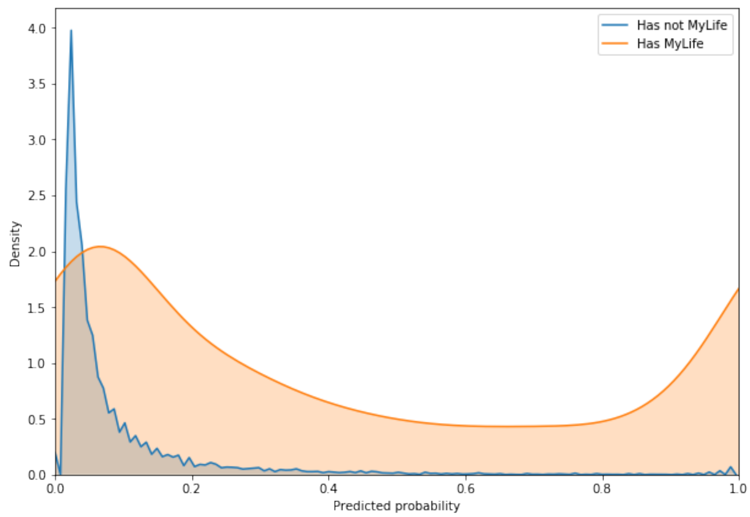

Figure 8: Density distributions for predicted probabilities for two classes.

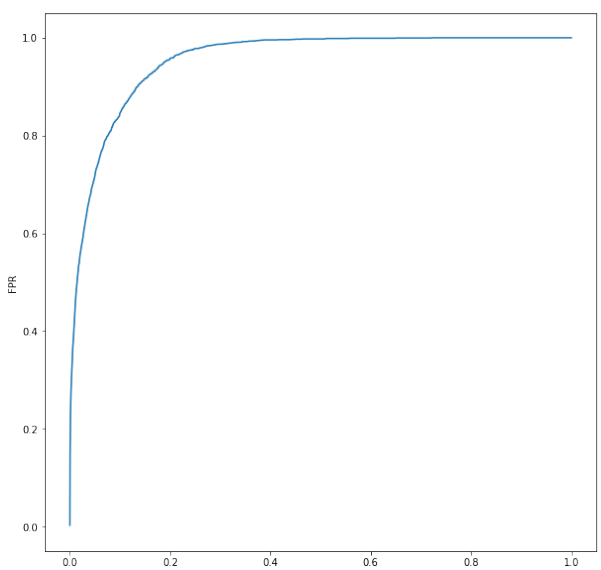

Figure 9: ROC curve for the classifier.

5.2.1 Feature analysis

Since the classifier performed well as discussed in Section 5.2, we can continue to feature analysis. Figure 10 shows shap values for every data point as well as indication of feature value itself. For a reminder, a shap value greater than 0 indicates that feature had a positive impact on the predicted probability a given customer having MyLife product and vice versa. From Figure 10 we can see that high investment activity (total income, salary, investable assets, housing loans and performing lending volumes) had a positive impact on the probability. In addition, high external assets (assets somewhere else than in Nordea) impacted the probability negatively.

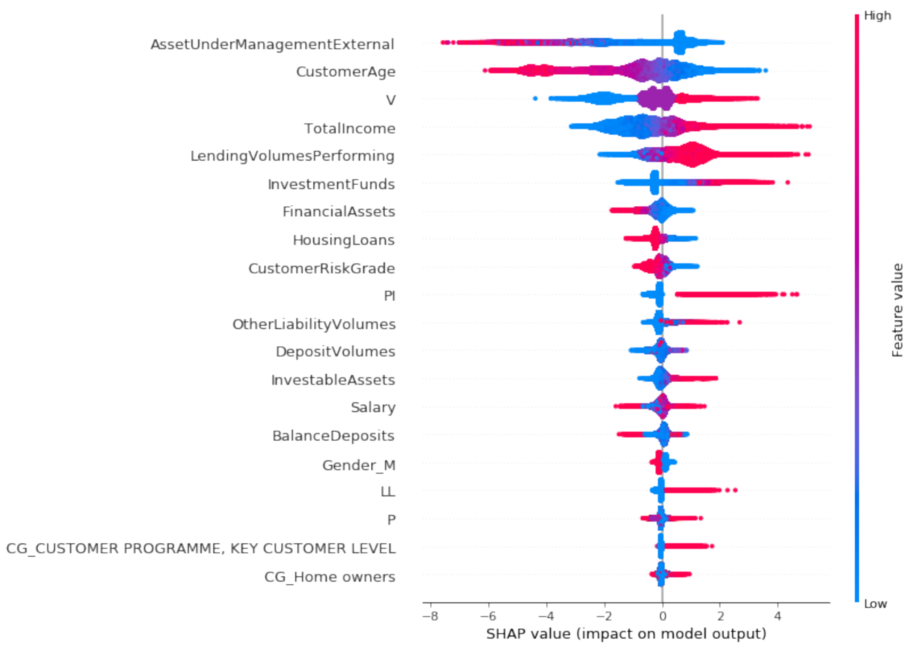

Figure 10: Shap values from the classifier.

By taking a closer look at the impact investment activity has on the probability of customer having MyLife product (Figure 11), we can see that the amount of total income is proportional to the prediction. In case of investment funds, shap values do not increase after a certain point, i.e., the positive impact on probability does not increase when the amount of investment funds surpases ≈ 500,000,000. Customer's who have a housing loan at Nordea seem to be more probable to have a MyLife product. In general, the more products customer had at Nordea, the greater the probability that a customer has a MyLife product (Figure 13). In our analysis, we only examined products/agreements with status "V", as we assumed this meant "Valid" and they represented the majority of product's statuses.

Moreover, customer's young age indicated greater probability of customer having MyLife product (Figure 12). This is in line with our assumption, that elderly people do not need life insurance products, since they are more expensive to them than for younger people.

From the feature analysis we are able to construct a target group, i.e., a group that is likely to buy a MyLife product. The customers Nordea Life Insurance should prioritize their marketing are young working people with existing services and products at Nordea. These types of customer's tend to have a relatively more MyLife products than their peers (Table 2).

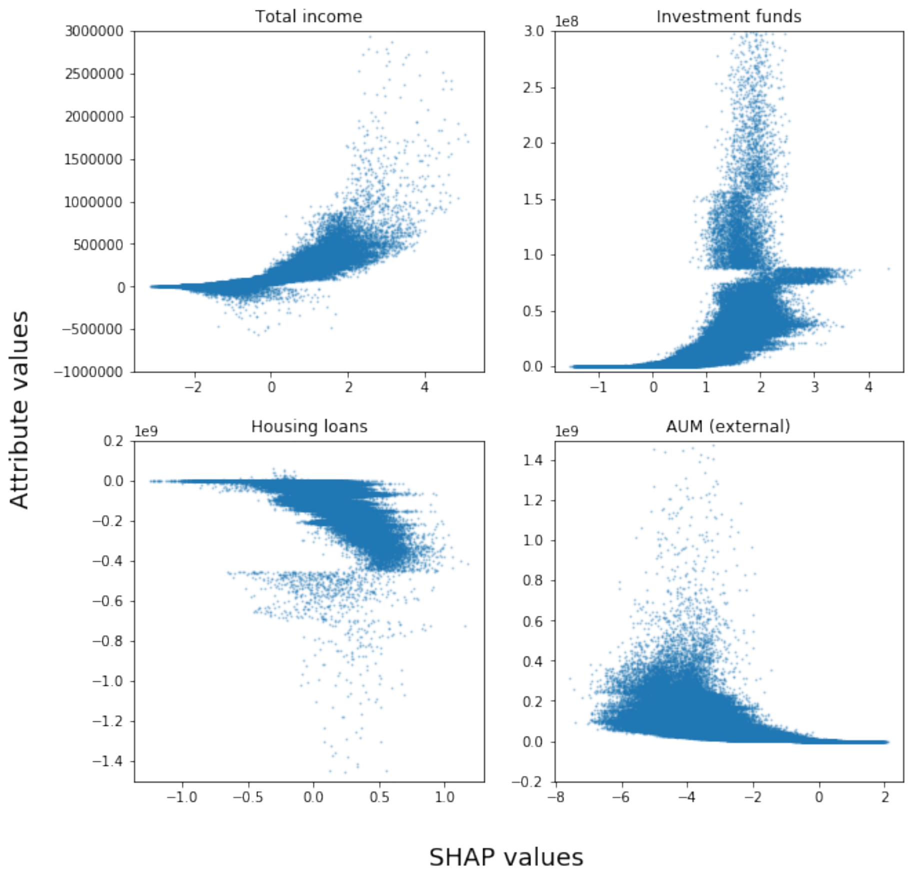

Figure 11: Dependency plots for investment activity.

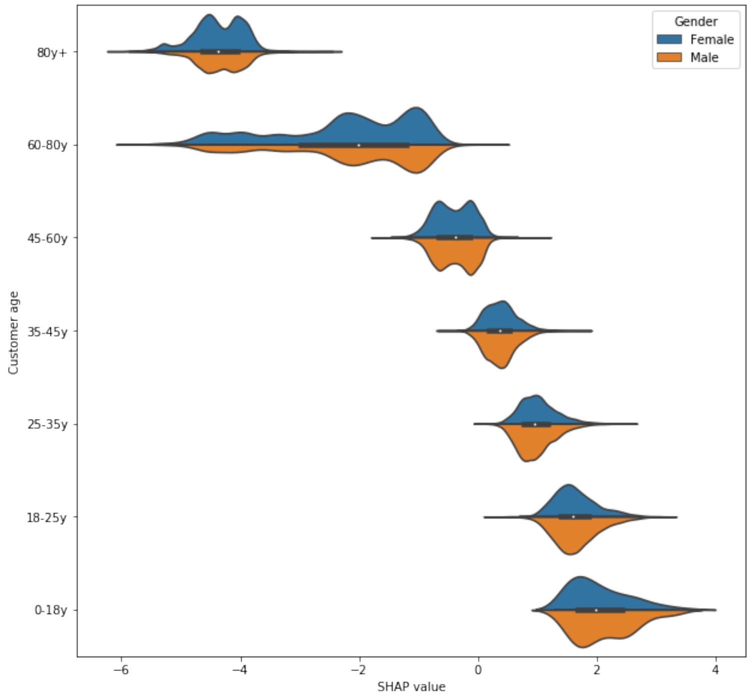

Figure 12: Dependency plots for customers' age.

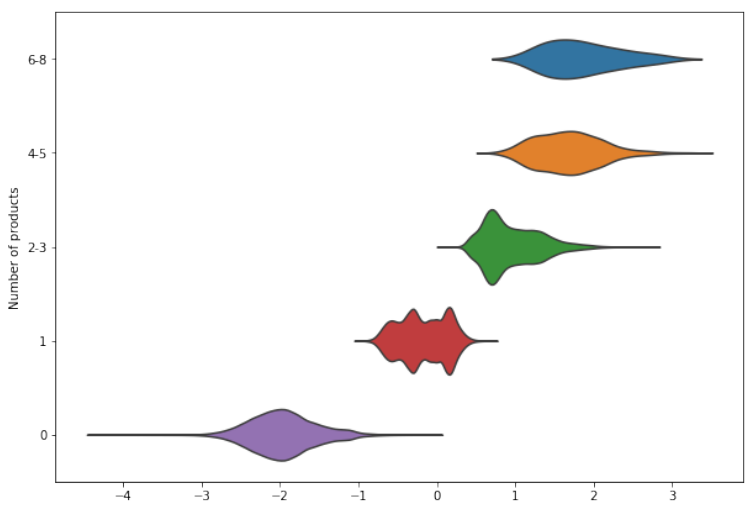

Figure 13: Dependency plot for number of existing products at Nordea.

Table 2: Percentage of customers who have MyLife per customers' age and total annual income.

CustomerAge	Total income				
	<18k	18-50k	50-100k	100k+	Total
0-18y	6.38	11.11	22.22	50	13.24
18-25y	6.06	13.66	16.18	29.98	13.18
25-35y	3.77	6.85	7.31	12.76	7.44
35-45y	1.28	2.04	2.93	7.33	3.23
45-60y	0.29	0.51	1.12	3.59	1.31
60-80y	0.04	0.12	0.22	0.49	0.21
80y+	0	0	0	0	0.00
Total	0.99	1.72	1.94	4.61	2.22

6 Discussion

6.1 Reflection on literature

The ideal model for CLV depends largely on the set purpose of CLV in context, or in other terms, how will the company plan to use the information and what kind of information is available. No single model offers all solutions, and data driven decision making should incorporate a collection of different models for different business decisions.

Often CLV is used for segmentation. Its possible that profitable customers do not represent any distinct group based on the data. Furthermore, it is important to understand that clustering is almost always a compromise: Choosing smaller clusters/segments allows more specific targeting, but makes efficient marketing customized for each segment process harder.

The literature review showed that there is considerable amount of information related to the CLV available. This is because eventually, CLV of current or future customers forms a proxy for firm value or its stock price. The similarity to the discounted cash flow, extensively used in finance, is striking. However, while many researchers have focused on developing theoretical frameworks for modelling CLV, many researchers have implemented a CLV model based on specific needs of a case project. Therefore, most of the information on CLV modelling available is highly tailored to a specific industry or client. This is understandable and expected, since CLV modelling is of high importance for most businesses.

The literature review also showed that many different types of CLV models have been developed. However, most approaches are defined and estimated at an individual customer or segment level that allows differentiation between customers. Most approaches also integrates customer retention rates into the CLV model.

This study shows an implementation of CLV modelling to the insurance industry. While the theoretical framework of this study is essentially the same as in literature, the implementation of the CLV model was fine-tuned to predict MyLife customers' CLV. This study adds to the literature on CLV modelling in retail banking.

6.2 Assessment of the results

One possible reason for the poor performance of the predictive model can be the lack of strong correlations in the data. The figure 14 shows different variables as a function of size of the monthly premium. Also the correlation coefficients are shown in figure titles. One can notice that the size of the premiums are not correlated with almost any of the variables. The strongest correlation (0.24) is detected between age and the size of the premium. This is due to the pricing of the insurances and one could expect even stronger correlation. There is no significant correlation between the other variables, even though one could expect for instance that persons having higher salary would be willing to buy larger life insurance with higher premium.

The correlations between the size of the premium and the other variables are crucial elements in the prediction of the CLV because the CLV model suggested in chapter 5.1 is strongly dependent on the size of the monthly premium. The lack of linear correlations implies that it might not be possible to predict the size of the premium based on data that was available to us. On the other hand, it could be possible that some more complex combination of the numerical and categorical variables would imply the CLV better.

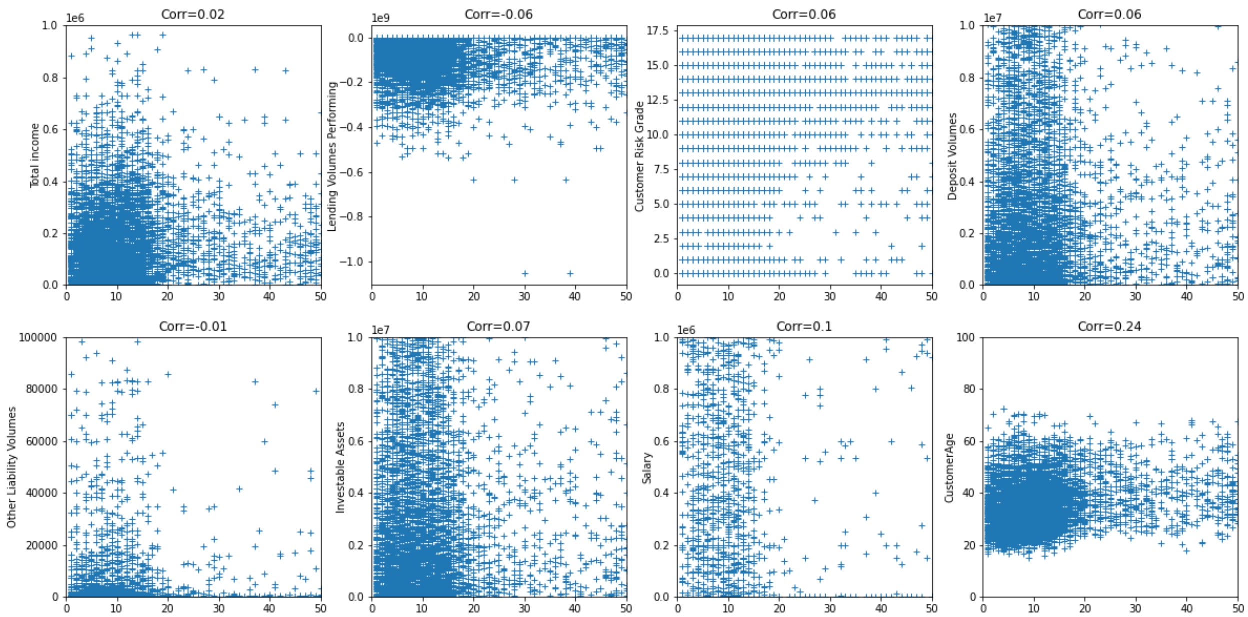

Figure 14: Different numerical data plotted as a function of monthly premium (the price of the life insurance). Correlations are shown in the titles.

What is the reason for the lack of correlations? One answer could be that the monthly premiums (and hence also CLV) is somewhat arbitrary and cannot be predicted from other customer features. It is possible that the price that a given customer wants to pay for his/hers insurance is affected by many things that are not visible in the data, such as the quality of the customer service or details in personal life.

Another, perhaps more probable, explanation could be that some of the correlations were lost in the process of simulating the data. As mentioned earlier, the data we used was not real customer data but it was simulated from the real data. This was done by adding noise to real data points. Adding only small noise to the data should not affect the correlations between features. However, if the added noise is larger and the correlations are not taken into account, some of the relationships in the data are lost in the simulation process. We clarify this with an example: if we add some random noise to persons salary, we should also increase the persons total income with a suitable amount. If this is not done the correlation between these two is weakened or even lost in the worst case. One possible reason for adding too large noise to the simulated data could be the numerous outliers in the data (customer age over 140, customer debt over 200 million euros, etc.). These outliers have enormous effect on the statistical measures of deviation and hence the amount of suitable noise might have been evaluated incorrectly.

7 Conclusions and strategy implications

Based on the literature review and the information gained in the process of exploring CLV models in practise we come up with the following final conclusions:

Analysing the current and potential customer monetary value in the context of Nordea risk insurance business requires sufficient information needed to estimate values for variables such as the retention rate and risk factors for existing and potential customers, which in turn implies a sufficient amount of high quality data is needed. If enough information is available, models such as presented in (4) can be used.

While the effect of simulation/anonymization is unknown, we concluded that the original dataset is likely to include a large number of outliers, inconsistenties and a large number of data is missing for a fairly large amount of customers. The current data management used in the case study example represents a typical database, or collection of databases that are designed or they have evolved to serve the traditional corporate business functions. This data system is not necessarily designed to serve the purpose of analytics. Current status quo requires query of data from multiple data sources.

Based on the former conclusions we propose the following strategic targets

- Customer data from different branches of the organization are mastered into a MDM platform according to the principles of Master Data Management (MDM) specifically for the purpose of analytics[23]
  - Meaningful data is available for analytics real time without querying different data sources
- Collecting, improving and maintaining customer data is set to be one of the key business functions of every branch of the corporation: Amount of quality customer data is considered to be an asset with future monetary value
  - Different business functions from direct sales to marketing and claim processing are designed to support data driven decision making at the top
  - Actively collecting key performance indicators, such as net promoter score, are included in the business processes at the B2C boundary
  - Existing data is appended with non-essential demographic data to aid future big data applications external sources
  - Long term goal of data collection is to provide sufficient information for estimating and predicting the complete lifetime cash flows on individual level
- Customer Lifetime Value Modeling is actively used in dictating the business decisions
  - The company aims to have real time CLV prediction of each individual customer
  - Example: Once a potentially churn triggering event has been detected, CLV estimation can be used to help decide whether the customer should be "let to churn" or whether the company should actively to prevent churn (such as by offering flexibility in payments in case where a sudden momentary loss of personal liquidity is detected.

8 Self assessment

The project successfully reviewed ways to calculate and model the customer lifetime value in the context of life insurance business. The suggested CLV value could not be tested and evaluated in detail due some missing data. The construction of predictive model for calculated CLV failed due to quality issues in the data.

The scope of the project remained almost the same during the project. Some discussion took place around whether we should focus more on determining customers CLV or finding potentially profitable customers without exact CLV value. Eventually, some time was dedicated to both of them. Delays in receiving the full data did not block us from carrying out he project as planned. However, it reduced the possibility of iterattively improving and rescoping the project.

In the beginning of the project, the risk of data quality issues was probably underestimated. Some of the problems might have been avoided if the problems were acknowledged in earlier stage. Also if the quality issues in the data would have been acknowledged earlier, there would have been more time to prepossess and clean the data. On the other hand, properly cleaning larger data files with tens of attributes would have taken a lot of time and still the result would have been uncertain. On the positive side, the possibility of difficult data was an important lesson learned and it is certainly worth remembering in future data science related projects.

The remote working due to the COVID-19 situation influenced the execution of the project. Although remote working sessions were arranged regularly and the team participated actively, it felt that the sessions were not as effective as some face to face sessions in the early spring. For instance, dividing work between the team members was not as smooth. The situation also affected the client organization and caused some delays in the process.

The team was asked to score the team and the project overally on scale from 1 to 5.

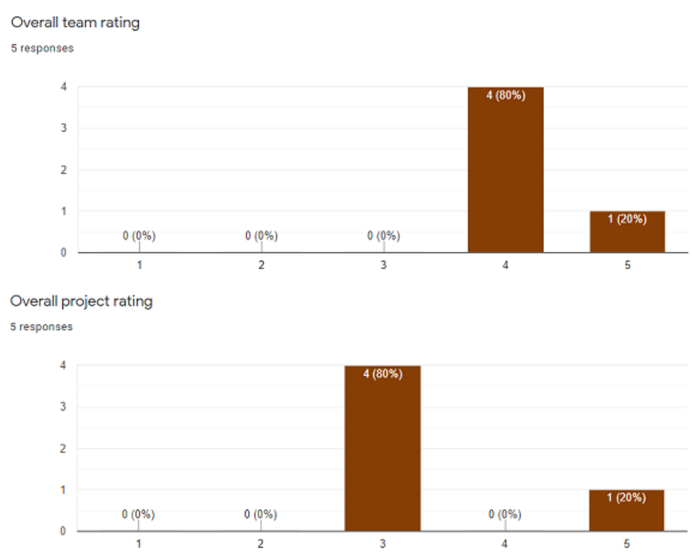

Figure 15: Overall project and team ratings

The team was also asked to score the teams strengths and weaknesses on scale where 1 implies weakness, 3 neutrality and 5 strength in that particular skill

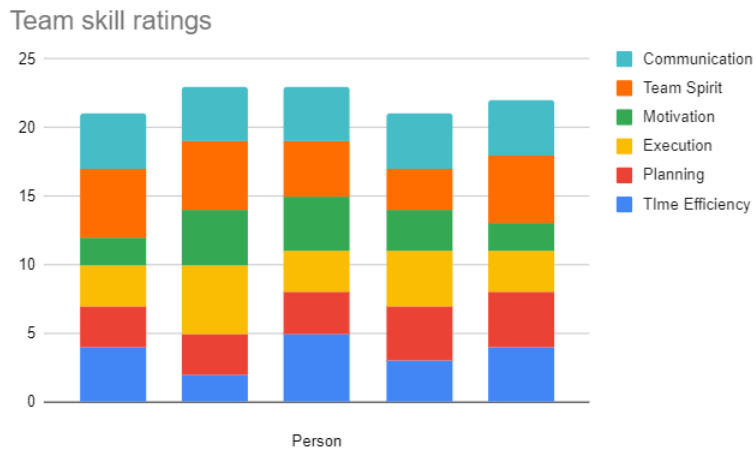

Figure 16: Team skill ratings

area. The cumulative scores stacked for each person are shown in Figure 16. Some indicators such as planning and time efficiency has some disparities, but the cumulative scores show that everybody seems to be overall satisfied with different areas of the team performance.

References

[1] Nordea, *Nordea at a glance*, [Online], `https://www.nordea.com/en/about-nordea/who-we-are/nordea-at-a-glance/` Cited 2.3.2020

[2] L. Thomas, *Customer Loyalty And Retention Primer*, [Online], `https://www.iiadallas.org/page/75`, cited 2.3.2020

[3] Nordea, *Fourth quarter and full year results 2019*, [Online], https://www.nordea.com/en/press-and-news/news-and-press-releases/press-releases/2020/02-06-06h30-fourth-quarter-and-full-year-results-2019.html, cited 2.3.2020

[4] K. Fang, Y. Jiang b, M. Song, *Customer profitability forecasting using Big Data analytics: A case study of the insurance industry*, Computers & Industrial Engineering 101 (2016) 554–564

[5] S. Gupta *et.al*, *Modeling Customer Lifetime Value*, Journal of Service Research, Volume 9, No. 2 (2006) 139-155

[6] V. Kumar, G. Ramani, T. Bohling, *Customer Lifetime Value Approaches and Best Practice Applications*, Journal of Interactive Marketing Volume 18, No. 3 (2004)

[7] T. Bayón, J. Gutsche, H. Bauer, *Customer equity marketing: touching the intangible*, European Management Journal 20(3) (2002) 213-222

[8] P. Berger, N. Nasr, *Customer lifetime value: marketing models and applications*, Journal of Interactive Marketing 12(1) (1998) 17-30

[9] S. Gupta, D. Lehmann, *Customers as Assets*, Journal of Interactive Marketing 17(1) (2003) 9-24

[10] H. Hwang, T. Jung, E. Suh, *An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunications industry*, Expert Systems with Applications 26 (2004) 181-188

[11] S. Gupta, D.R. Lehmann, J.A. Stuart, *Valuing customers*, Journal of marketing research, 41(1) (2004) 7-18.

[12] N. Glady, B. Baesens, C. Croux, *Modeling Churn Using Customer Lifetime Value*, European Journal of Operational Research, 197(1) (2009) 402-411

[13] S. Kim, T. Jung, E. Suh, H. Hwang, *Customer segmentation and strategy development based on customer lifetime value: A case study*, Expert Systems with Applications 31 (2006) 101–107

[14] T. Mutanen, J. Ahola, S. Nousiainen, *Customer churn prediction - a case study in retail banking*, Proc. of ECML/PKDD Workshop on Practical Data Mining (2006) 13-19

[15] R. Blattberg, J. Deighton, *Manage Marketing by the Customer Equity test*, Harvard Business Review (1996) 136-144

[16] W. Reinartz, V. Kumar, *On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing*, Journal of Marketing, 64 (2000) 17-35

[17] J. Thomas, *A Methodology for Linking Customer Acquisition to Customer Retention*, Journal of Marketing Research, 38(2) (2001) 262-68

[18] P. Fader, G. Bruce, K. Lee, *RFM and CLV: Using Iso- CLV Curves for Customer Base Analysis*, Journal of Marketing Research, 42 (2005) 415-30

[19] Lifelines documentation. [Online]. `https://lifelines.readthedocs.io/en/latest/`

[20] M. Goel, P. Khanna, J. Kishore, *Understanding survival analysis: Kaplan-Meier estimate*, International Journal of Ayurveda Research (2010) 274-278

[21] Y. LeCun, Y. Bengio, G. Hinton, *Deep learning*, Nature, 521 (2015) 436–444

[22] T. Hastie, R. Tibshirani, J.H. Friedman, *10. Boosting and Additive Trees*, The Elements of Statistical Learning (2nd ed.), New York: Springer 337–384

[23] Kumar, Pavan, *Master Data Management (MDM) – Strategies, Architecture and Synchronisation Techniques*, (2014), 10.13140/RG.2.1.1379.9846.

[24] Helsinki: Statistics Finland, *Official Statistics of Finland (OSF): Deaths*, `http://www.stat.fi/til/kuol/index_en.html`, cited 12.5.2020.

[25] S. M. Lundberg *et.al,*, *From local explanations to global understanding with explainable AI for trees*, Nature Machine Intelligence 2 (2020) 56-67

[26] Owen, Richard, *Net Promoter Score and Its Successful Application* (2019) 10.1007/978-981-10-7724-1_2

Appendices

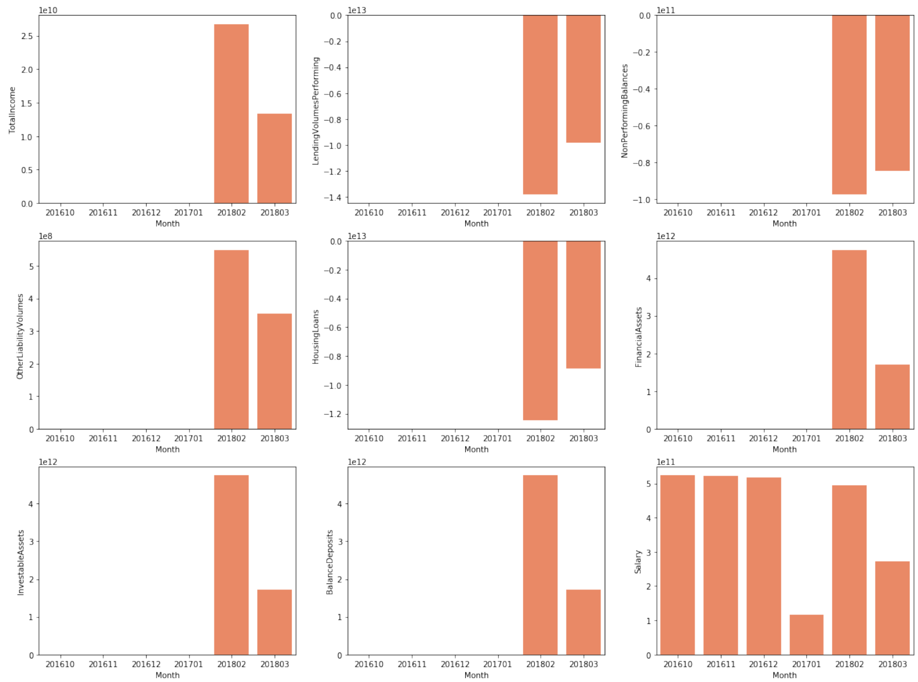

Figure 17: Total sum of attribute values for time series data attributes per month from a relevant subset.

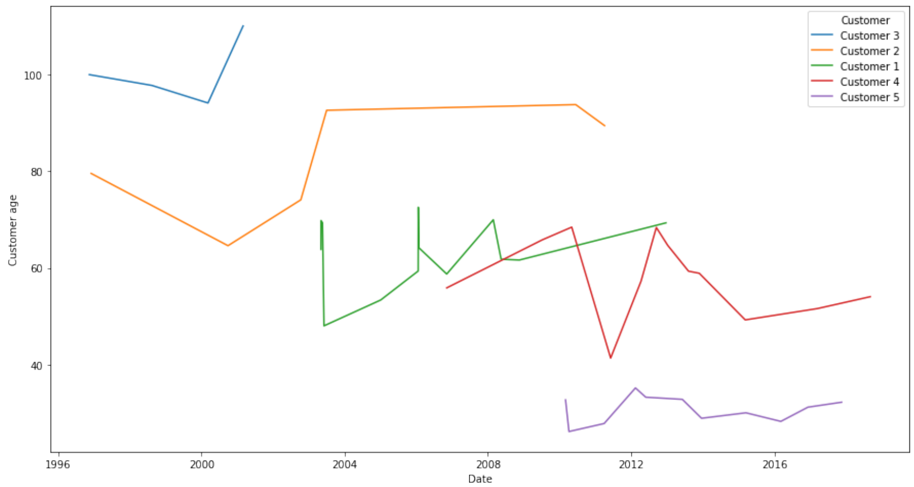

Figure 18: Age for 5 different random customers.