11-28-2018

# FEASIBILITY OF B2C CUSTOMER RELATIONSHIP ANALYTICS IN THE B2B INDUSTRIAL CONTEXT

Kati Stormi
*Tampere University of Technology*, kati.stormi@gmail.com

Teemu Laine
*Tampere University of Technology*, teemu.j.laine@tut.fi

Tapio Elomaa
*Tampere University of Technology*, tapio.elomaa@tut.fi

# FEASIBILITY OF B2C CUSTOMER RELATIONSHIP ANALYTICS IN THE B2B INDUSTRIAL CONTEXT

*Research paper*

Stormi, Kati, Tampere University of Technology, Finland, kati.stormi@gmail.com

Laine, Teemu, Tampere University of Technology, Finland, teemu.j.laine@tut.fi

Elomaa, Tapio, Tampere University of Technology, Finland, tapio.elomaa@tut.fi

## Abstract

*The purpose of the paper is to evaluate the feasibility of business-to-consumer (B2C) customer relationship analytics in the industrial business-to-business (B2B) context, in particular spare part sales. The contribution of the paper is twofold; the article identifies analytics approaches with value potential for B2B decision-making, and illustrates their value in use. The identified analytics approaches, customer segmentation, market basket analysis and target customer selection, are common in the B2C marketing and e-commerce. However, in the industrial B2B marketing, the application of these approaches is not yet common.. The different kinds of analytics under examination in this paper use machine learning (ML) techniques. The examination takes into account the applicability and usefulness of the techniques as well as implementation challenges. The research suggests that the identified analytics may serve different business purposes and may be relatively straightforward to implement. This requires careful examination of the desired purposes of use in a particular business context. However, the continuous and real-time use of such analyses remains a challenge for further examination also in information systems research.*

*Keywords: Business analytics, B2B decision-making, Machine learning, Data mining, Artificial intelligence, CRM*

# 1 Introduction

Increased awareness of the customer relationships represents a starting point for competitive advantage. Every transaction is an opportunity for the company to collect information about the customer. Sales history shows the customer's product preferences, purchase volumes and frequency. Aggregating purchase history over all customers gives information, for example, about the typical market basket, i.e. what products customers usually buy together. For example, one may find out that 70 percent of customers who bought a red dress also bought red high heels. In addition to sales information, companies can collect many other pieces of information about their customers. In the business-to-consumer (B2C) marketing, companies can use customer demographics, such as age, gender, income, geographical location and occupation, to segment their customer base. In the business-to-business (B2B) context, customer firmographics, such as industry, geographical location, size and performance serve the same purpose (Weinstein, 2014). In all, customer information is an asset that can enhance customer acquisition and retention as well as cross-selling and upselling. In addition, analyzing customer information can help companies improve customer satisfaction; customer information may support the identification and fulfillment of customer needs in different ways.

The amount of customer sales data available can be tremendous. For example, biggest e-commerce companies can handle millions of orders daily. Consequently, companies need computing power, advanced databases, artificial intelligence (AI) algorithms, and different suitable tools, to process, store and analyze customer sales information. Companies can use sales information for a variety of business purposes with different techniques. Lin *et al.* (2017) identifies eight business application areas such as customer relationship management (CRM), recommender systems and bankruptcy prediction. Analytics can support multiple dimensions of CRM: customer identification, attraction, retention and development (Ngai *et al.*, 2009). For example, companies can use different clustering algorithms to segment their markets (e.g. Sarvari *et al.*, 2016; Olson and Chae, 2012), to target marketing (e.g. Gordini and Veglio, 2015), to forecast sales and customer lifetime value (e.g., Stormi *et al.*, 2018), to do churn prediction (e.g. Gordini and Veglio, 2017), to make market basket analysis (Ramya and Ramakrishnan, 2016), and to create recommended systems (e.g., Geuens *et al.*, 2018).

In the B2B settings, companies search for profitable growth from the existing customer relationships (Anderson and Narus, 2003), and the companies need to be increasingly aware of the value potential embedded to their customer relationships. Despite the identified need for managing the mutual value of the customer relationships in the B2B settings (see e.g., Selos *et al.* 2013), it seems that customer relationship analytics are more common in the B2C than in the B2B marketing (Martínez-López and Casillas, 2013) and many B2B companies are struggling with data gathering and analysis (Cortez and Johnston, 2017). In addition, advanced analytics typically require skills in mathematics, statistics and programming that accounting and marketing professionals do not necessarily have, and accountants and marketers do not necessarily know what and how much data are needed and how to filter information (Cortez and Johnston, 2017). Consequently, companies need data scientists to complement the tacit knowledge of accounting specialists and marketers (Al-Htaybat *et al.*, 2017), in order to conduct customer relationship analyses.

In response to the identified research gap on the potential of advanced analytics in examining the value of B2B customer relationships, this paper focuses on examining the customer relationship analytics potential in the industrial B2B settings. Thus, *the purpose of the study is to evaluate the feasibility of B2C customer relationship analytics in the industrial B2B context.* In such evaluation, the study uses both supervised (classification) and unsupervised machine learning (ML) algorithms (clustering and association rules). The analyses presented in this paper are descriptive and predictive, and they use structured data on spare parts sales and industrial customers stemming from a real-life case. The evaluation of the results is based on detailed documentation and management judgement. In addition, to more comprehensively address the feasibility of the customer relationship analytics the evaluation considers also implementation challenges confronted during the examination.

The contribution of the paper is twofold; the article identifies analytics approaches with value potential for B2B decision-making and illustrates their value in use. These contributions respond both to the need for understanding the value dynamics of the customer relationships (Anderson and Narus, 2003; Selos *et al.*, 2013), and to the recently identified lack of B2B applications of advanced customer relationship analytics in the B2B settings (Martínez-López and Casillas, 2013; Cortez and Johnston, 2017). Furthermore, the paper contributes on the industrial B2B marketing literature by reflecting upon different customer relationship analytics in the given context especially by building at Stormi *et al.* (2018) and Dekker *et al.* (2013). In addition, the paper contributes on literature about advanced business analytics, AI and ML in particular, in the industrial B2B marketing management. As a practical implication, the study aims to help industrial companies that are collecting diverse data about sales and customers. However, the companies do not necessarily know what kinds of analytics to apply on such data, and they do not necessarily have competence to carry out the analysis. In practice, companies wish to get deeper insights on data to supplement descriptive analytics on past sales and profitability, but they do not know exactly how. The study reflects upon what kind analyses companies could apply and what kind of tools and capabilities they need for the implementation. To unveil practical restrictions, the paper reflects upon the time spent on analyses under examination.

Empirically, the research is a single case study. The case company, called ScrewCo for confidentiality, is a large original equipment manufacturer (OEM) that manufacturers and serves large mobile industrial equipment globally. ScrewCo collects and maintains a comprehensive database about spare parts sales and the global installed base. The sales database covers all spare parts orders from the beginning of the 2010s. Regarding the installed base data, the oldest pieces of equipment recorded in the database are already from 1970s. ScrewCo has a comprehensive analytic department that constantly monitors sales and profitability of the company. However, the company is lacking deeper insight on the dynamics of the customer relationships and customers' buying behavior. In addition, ScrewCo wants to identify customers whose purchasing behavior is unusual compared to other similar customers.

The rest of the paper is organized as follows. Chapter 2 discusses the literature on B2C customer relationship analytics and introduces industrial B2B marketing management and its specific features compared to B2C marketing management. Chapter 3 presents the used research method, the case company and the used data. Chapter 4 introduces the analytics conducted and their results. Finally, Chapter 5 concludes the implications of the research findings.

# 2 Literature review

## 2.1 CRM and business analytics

Customer relationship management (CRM) refers to all actions that companies take to enhance existing or future customer relationships to drive sales and profitability growth (Kumar, 2010). CRM covers people, processes and technology used for enhancing customer satisfaction and company performance (Chen and Popovich, 2003). CRM technologies, within the wider scope of customer relationship management, are specialized in collecting, storing and analyzing customer information. Companies can collect (at least) three types of information about their customers: customer personal data, action data and reaction data (Reimer and Becker, 2015). *Customer personal data* refers to customer demographics (e.g. age, gender, geographical location, income, occupation and education) in the B2C marketing and to firmographics (e.g. industry, geographical location, size and performance) in the B2B marketing. *Action data* refers to actions that company has made to enhance customer relationship, such as phone and email contacts, special offers, personal selling and campaigns. *Reaction data* describes the reaction of the customer to the actions of the company. Ultimately, reaction data comprises customer sales information: sales volume, frequency and purchased products.

The amount of customer data can be considerable. Therefore, companies often use advanced analytics techniques to process the data. Advanced marketing analytics discover patterns in data, provide deep insights and knowledge on data and use intelligent techniques (e.g. AI, ML, neural networks and text/video/audio/image mining methods) and statistical techniques (e.g. linear and logistic regression and Naïve Bayes) to analyze the data (Lin *et al.*, 2017). The orientation of advanced customer relationship analytics can be descriptive (what has happened), predictive (what will happen) of prescriptive (what should be done) (Appelbaum *et al.*, 2017). The used data can be structured or unstructured. *Structured data* is typically organized into rows (items) and columns (data attributes) and it is stored in a database. Meanwhile, the semantics of *unstructured data* are not necessarily explicitly stated. Unstructured data cannot be possessed and analyzed easily using most database management systems and software. Typically, unstructured data is text, such as email messages or textual social media content, but it can also contain for example audio, video and image files (Warren Jr *et al.*, 2015). In addition, the volume of the unstructured data can be huge. For example, unstructured data can cover all product reviews of online customers of the company. Consequently, *big data* is closely related to the concept of unstructured data. Syed *et al.* (2013) estimate that about 90 percent of big data is unstructured. Three V's define big data: Volume, Variety and Velocity (Gandomi and Haider, 2015). Volume refers to the magnitude of the data, that can be terabytes or even petabytes, variety refers to structural the structural heterogeneity of the data, and velocity refers to the speed of data creation (Gandomi and Haider, 2015).

Prior to applying advanced analytics, the company needs to answer two important questions. First, the company has to decide, what will be analyzed, i.e. what is the purpose of the analysis. Second, the company has to decide, what kinds of data to collect and what kinds of techniques and tools to apply, i.e., the decisions on the actual implementation of the analysis. To get a broad picture of the first problem, Lin *et al.* (2017) recently identified eight business application areas for advanced business analytics: Bankruptcy prediction, CRM, fraud detection, intrusion detection, recommender systems, software development effort estimation, stock prediction and other financial time-series areas. Also recently, Appelbaum *et al.* (2017) used four perspectives of the balanced scorecard (BSC) (financial, customer, internal business processes and learning and growth) for the classification of business analytics application areas. Bose and Mahapatra (2001) used functional area for the classification: finance, telecom, marketing, web analysis and others. In addition, companies can use advanced analytics, for example, in corporate reporting (Al-Htaybat *et al.*, 2017), to enhance industrial system reliability and security and to manage operational risk (Choi *et al.*, 2017), to predict financial performance and stock markets (Fisher *et al.*, 2016), and for supply chain management (Trkman *et al.*, 2010) and performance management (Schläfke *et al.*, 2012) in general.

From the perspective of this study, advanced business applications in CRM are particularly interesting. In order to add value to the B2B decision-making, advanced business analytics can support all four dimensions of CRM: customer identification (target customer selection and customer segmentation), customer attraction (direct marketing), customer retention (customer profiling, recommender systems) and customer development (customer lifetime value analysis and market basket analysis) (Ngai *et al.*, 2009). Similarly, Noori and Salimi (2005) identify three application areas: customer profiling, deviation analysis and trend analysis over time. In the industrial marketing context, Martínez-López and Casillas (2013) mention, for example, segmenting and targeting business markets, managing customers' relationships and web intelligence and B2B e-commerce applications. In all, in particular customer segmentation, market basket analysis, target customer selection and recommender systems are already popular forms of customer relationship analytics. Next, customer segmentation, market basket analysis and target customer selection are discussed in more detail.

## 2.2 Clustering for customer segmentation

Clustering means organizing items into groups, whose members are sufficiently similar to the clustering variables. In the CRM context, companies typically use clustering to segment their customers. Customer segmentation, thus, groups similar customers together and separates dissimilar ones (Bose and Chen, 2009). In the B2C marketing, companies can use variety of clustering variables, such as consumer demographics (gender, age, location, education and income). Correspondingly, market segmentation can be based on firmographics (e.g. industry, location, size and performance) in the B2B marketing. In addition, companies can segment customers based on customer purchase history, i.e. what, how much, where, when, how often customer has bought (Reimer and Becker, 2015). Recency, frequency and monetary value (RFM) variables summarizes a customer's purchasing behavior characteristics: time of most recent purchase (recency), number of transactions (frequency) and the average or total values of purchases (monetary value) in the past (Fader *et al.*, 2005).

RFM analysis is a prominent example of simple management heuristic (Wübben and Wangenheim, 2008) for customer segmentation. RFM analysis divides customers into profitability segments based on simple rules on their recency, frequency and monetary value characteristics. For example, the best customers have bought recently (e.g., less than one month ago), they buy often and purchase a lot (e.g. over 10 transactions and € 5,000 during in the last two months). Besides simple heuristics, companies can use different machine learning algorithms for segmentation. Analysts decide which variables to use for segmentation (e.g. RFM variables) while algorithms make the trick, i.e., find suitable customer segments. Sarvari *et al.* (2016) found that combining RFM variables with demographic variables produce segments that are more accurate.

Clustering is an example of an unsupervised machine learning approach. Unsupervised ML algorithms identify patterns in data and draw inferences from data without the right answers. There are abundant clustering algorithms. Choosing the right algorithm is not trivial and it requires experience and understanding of the algorithms (Nairn and Bottomley, 2003). In particular, the number and shape of clusters influence the choice of the right algorithm. K-Means is a general, simple and widely used clustering algorithm across different disciplines (Wu *et al.*, 2008). Affinity propagation and DBSCAN are examples of other clustering algorithms. The challenge of applying the K-Means algorithm is that a user must pre-choose the number of clusters. There are several different techniques for choosing the right number of clusters, such as distance between data points, cross-validation, information criteria and silhouette method. Affinity propagation, in turn, chooses the number of clusters based on the data. However, the algorithm is complex and depending on the data, and the execution time can be long. Applications of customer segmentation in the marketing context remain abundant (e.g. Hosseini *et al.*, 2010; Chen *et al.*, 2012; Güçdemir and Selim, 2015; Sarvari *et al.*, 2016; Dursun and Caber, 2016)

## 2.3 Association rule analysis for market basket analysis

Market basket analysis identifies combinations of products that customers frequently purchase together. Thus, it identifies relationships between groups of products, items, or categories (Aguinis *et al.*, 2013). Companies can use market basket analysis to identify complementary items (Chen *et al.*, 2005). Changes in the price of an item also affect the demand of the complementary product. In addition, market basket analysis can be used in deciding the location of items inside a store (Russell and Petersen, 2000) also many recommendation systems are based on association rule mining (Gatzioura and Sànchez-Marrè, 2015).

Association rule analysis is the ML technique to uncover how products are associated to each other. There are several algorithms for determining association rules (e.g. Apriori or Eclat). The algorithms produce a set of association rules, for example {tomato, cucumber} → {salad}, meaning that customers that bought tomatoes and a cucumber are also likely to buy salad. There are several different indicators describing the strength of the association. *Support* is the percentage of transactions over all transactions that contain the given association. *Confidence* is another measure for the strength of the association. In the example given, the confidence is $\frac{Supp(tomato,cucumber,salad)}{Supp(tomato,cucumber)}$, i.e., percentage of transactions including salad over all transactions including tomatoes and cucumber.

## 2.4 Classification for selecting target customers

Target customer is a customer that is most likely to buy your product or products. Therefore, companies should focus product marketing for the target customers. Companies have many ways to identify target customers. Quite often, companies first segment their customers (based on e.g. customer demographics, firmographics, preferences and buying history) and then market products to customer segments that are most likely to buy products. In addition, companies can use ML algorithms, classification in particular, to identify target customers, i.e., customers that are likely to buy a product (e.g., Jiang *et al.*, 2012; Qiu *et al.*, 2015; Pilerood and Gholamian, 2016). Generally speaking, classification is a ML technique that assigns categories to a collection of items. For example, banks can classify their customers into two categories: default and non-default customers. Default customers are likely to default whereas non-default customer are likely to repay the loan immediately. The result of the target customer selection may be binary: the customer either buys or does not buy the product. In addition, classification algorithms can predict the likelihood that a customer purchases a product (Jiang *et al.*, 2012).

## 2.5 Industrial B2B marketing management as context

In the paper, industrial B2B marketing management represents the business context, in which the feasibility of the different forms of analytics are examined. In this paper, industrial B2B marketing refers to business transactions that involve the sale of one company's product or services to another industrial company. Typically, the sold product or service is an integral part of customer's own industrial production, ranging from complicated product lines to single industrial equipment. The sales consist typically of original equipment sales and after-sales such as maintenance, repair and spare parts. In the industrial B2B context, pieces of equipment are typically large and expensive.

The is examination is due to the so far insufficient understanding about the customer behavior within the B2B relationships; in particular regarding the current and potential value of those customer relationships (Stormi *et al.* 2018, see also Anderson and Narus, 2003). Indeed, if the B2B customer behavior is thoroughly understood, the value of the customer relationships can be managed. In practice, this can mean, for example, anticipating customers' buying behavior (Stormi *et al.*, 2018), estimating the life-time potential of new business opportunities (Lindholm *et al.*, 2017), and preventing unwanted shifts of the existing customers to other providers (Selos *et al.*, 2013). In all, the potential of applying the advanced analytics lies in the possibilities to unveil the dynamics of the value of the B2B relationships and extended abilities to manage such dynamics.

B2B context embeds some characteristics that affect the use of the advanced analytics in this context. Compared to the B2C marketing, the number of customers is relatively low, for example some thousands of customers (Noori and Salimi, 2005), and the markets are characterized by low volume but high value transactions (Berthon *et al.*, 2003). In addition, selling in the B2B context is more often relationship-based than transaction-based (Oliva and Kallenberg, 2003). Further, relationship satisfaction, trust and commitment increase customer loyalty and satisfaction (Chumpitaz Caceres and Paparoidamis, 2007). Consequently, for example the online stores are not as common as in the B2C marketing. In the industrial B2B context, installed base of equipment in use by customer is a central concept. The installed base includes all the pieces of equipment sold by the original equipment manufacturer (OEM) to its customers that will be served by the OEM and other service providers during the equipment's lifecycle (Dekker *et al.*, 2013; Oliva and Kallenberg, 2003). On one hand, B2B context provides new data sources, regarding the installed base and customers' business operations, for instance, for understanding the value of the B2B relationships. On the other hand, the feasibility of the advanced analytics in managing the B2B relationships is largely unknown in theory and in practice, despite the seemingly large potential inherent to it.

# 3 Method and data

Empirically, this study is based on a single case study that illustrates the feasibility of the advanced analytics in managing the B2B customer relationships. Case study fits well to the purpose of the study: to test the feasibility of customer relationship analytics in the industrial B2B context. Thus, the study examines and evaluates (but does not prove) the applicability and prevalence of the selected analytics techniques in the industrial B2B context.

Essentially, the purpose of the study has also directed the choice of the case company. The case company must be an industrial service organization that collects appropriate business data, but does not necessarily apply advanced analytics on data. Although the case study results cannot be directly generalized, it is fair to assume that many industrial service companies collect similar data about their sales and customers and might benefit from the results of the study in their endeavor to improve customer relationship analytics. Besides, given the novelty of the particular research area, there are not many other studies examining the feasibility of similar approaches in similar (or related) business contexts.

Case company, called ScrewCo for confidentiality, is a large OEM that sells and serves large industrial equipment globally. The installed base covers thousands of equipment with an operational life cycle of typically tens of years. The share of industrial services in the revenues of ScrewCo is over 50 %. Services cover maintenance, repair and spare parts. ScrewCo has thousands of customers globally. The size of customers' installed base varies greatly: for some customers, the installed base includes hundreds of pieces of equipment whereas many customers have only one piece of equipment.

ScrewCo collects and maintains a comprehensive database about spare part sales and installed base. Since the beginning of the 2010s, ScrewCo has tracked all spare part orders: sold spare parts and their material, order size (in euros and number of articles) and customer identification. In addition, ScrewCo maintains a database about installed base: the date sold, equipment type and original customer. ScrewCo applies descriptive analyses and visualization on data, such as sales and profitability by product, customer and market area. ScrewCo segments customers based on industry and geographical location. The segmentation is currently based on simple heuristic and the company does not use, for example, clustering algorithms for categorizing customers. In short, ScrewCo collects comprehensive information about installed base and sales. However, especially customer relationship analytics is still at the forefront.

ScrewCo has identified new requirements for the customer relationship reporting and analytics. First, analyses should be more predictive. Better sales forecasts would support service production and spare part inventory management. Second, analyses should pinpoint opportunities to increase sales. Third, analyses should increase information about customers and their purchasing patterns, for example, analyses should help the company identify lost customers or customers that ScrewCo is in danger of losing. In addition, analyses should help the company identify customers that consume spare parts abnormally compared to other customers with similar installed base.

The research process underlying the paper took the following steps. The researchers received the data about spare part sales and customers' installed base in the spreadsheet form in February 2017. The analysis phase started in March 2017. During the analysis phase, the researchers kept a diary about conducted tasks and time spent on the analyses. The researchers had an earlier programming experience. However, ML tools and algorithms were not familiar. After the analyses were completed, the results were presented to the management in June 2017. The analysis phase started in March 2017 and the results were presented to company managers in June 2017. In total, the research data consists of three main elements: Quantitative data about spare part sales and customer's installed base, research diary and interview about the results of the analyses (Table 1). The next section presents the analyses made, the used data and the main results.

| ***Data*** | ***Type*** | ***Description*** |
|---|---|---|
| selected spare part sales to selected customers over 2011 - 2016 | Quantitative | Structured data initially in the spreadsheet form |
| Customer installed base information | Quantitative | Structured data initially in the spreadsheet form |
| Research diary | Qualitative | A diary about analyses made and the time spent threin |
| Interview about analyses results | Qualitative | |

*Table 1. The research data*

# 4 Case study

## 4.1 The selection of advanced business analytics techniques

The first task in the data analyses was to select suitable techniques to apply on the data. Researchers and the company management together decided to apply two widely used marketing analytics on data: customer segmentation and market basket analysis. In addition, they decided to use classification algorithms to identify active customers, i.e., customers that are likely to buy maintenance services next year. Customer segmentation would give information about typical buying patterns of customers. For example, some customers might prefer several small orders whereas others can make big orders infrequently. In addition, sales volumes against the size of the installed base of the customer may vary: some customers buy a substantial amount of spare parts in relation to their installed base, whereas others buy less than expected. Market basket analysis identifies typical spare parts that customers buy together. This information will help ScrewCo to serve customers better by recommending other spare parts that are likely to wear at the same rate. Customer classification can serve many purposes. In this study, classification was used to identify active customer, i.e., customers that are likely to make an order in the coming year. This information is vital since ScrewCo has several customers who do not make orders yearly. Identification of prominent active customers helps ScrewCo to target sales more effectively. Next, the conducted analyses are described in more detail.

## 4.2 Customer segmentation

The data covered spare parts sales in a selected market area from 2011 to 2016 and customer installed base information. Originally, the number of customers was 1 049. However, those customers who had not made orders for two consecutive years were considered lost and they were left out of the analysis. In addition, biggest customers were excluded from the analysis. Consequently, the segmentation comprised of 751 customers. The used segmentation variables were average yearly purchase frequency (F), monetary value from 2011 to 2016 (M) and the size of the installed base in 2016 (IB). F, M and IB variables were chosen as sales volumes and installed base size of customers vary considerably. In addition, ScrewCo wanted to get better understanding about customer buying behavior, for example some customers might prefer several small orders while others make bigger orders less frequently. Further, some customers order considerably less compared to the size of their installed base.

Scikit-learn machine learning library for Python and two different clustering algorithms were used: K-means and Affinity propagation. K-means algorithm requires the number of clusters to be specified whereas Affinity propagation chooses the number of clusters based on the data. Figure 1 shows results for the K-means algorithm with four clusters. The figure depicts clusters in a two-dimensional space: average purchase frequency against monetary value in 2011 – 2016. For confidentiality reason, the figure does not reveal exact numbers. Instead, F and M variables are divided into size categories XXS - XXL. Overall, Figure 1 shows that in the two dimensional-space (F and M variables) customers do not form clear clusters. Instead, F and M variables correlate strongly, i.e., the higher the monetary value the more often the customer makes orders.

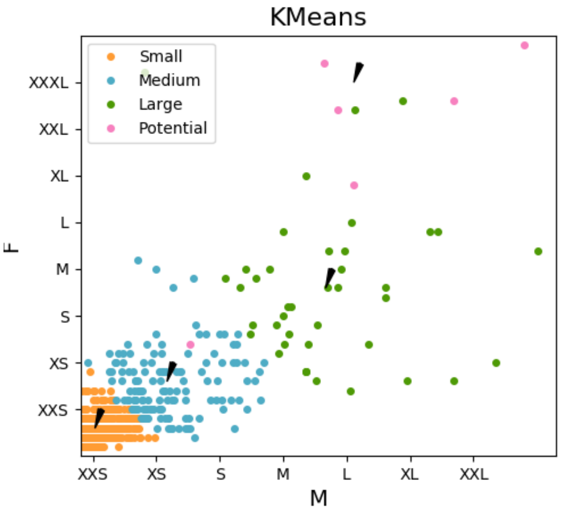


*Figure 1. The clustering results for the K-Means algorithm*

In Figure 1, black arrows point to the cluster centers. In addition, Table 1 shows cluster center values for all clustering variables, including the size of customer's installed base. In the first three segments, Small, Medium and Large, all three variables (F, M and IB) are rising steadily. However, the Potential segment is different. The monetary value of Potential customers is almost the same than Large customers, yet the F variable, and especially the IB variable, are multiple compared to the Large customers. This means that Potential customers buy little spare parts in relation to the size of their installed base. In addition, their purchase frequency is relatively high compared to the Large customers. Consequently, ScrewCo has the opportunity to increase spare parts sales to the Potential customers.

| ***Cluster*** | ***M*** | ***F*** | ***IB*** |
|---|---|---|---|
| **Small** | XXS | XXS | XXS |
| **Medium** | XS | XS | XS |
| **Large** | M | S | M |
| **Potential** | M | XXL | XXXL |

*Table 2. Cluster centers*

Figure 2 shows the segmentation results for the Affinity propagation algorithm.

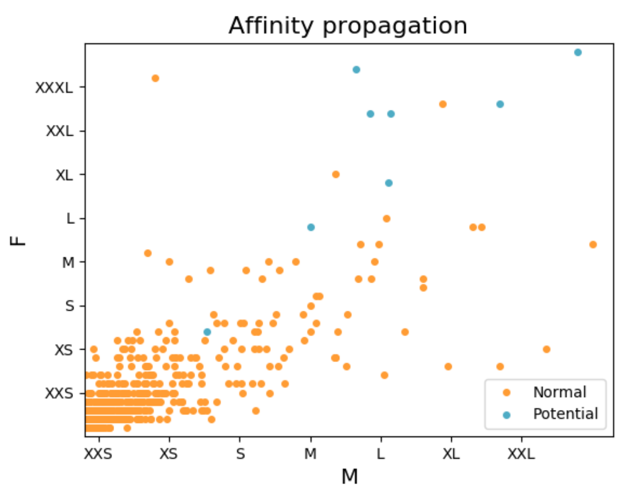


*Figure 2. The affinity propagation algorithm clusters*

Affinity propagation chooses the number of cluster based on the data. With the given preference (-200), the algorithm identified two customer segments: segment one (Normal) includes customers whose monetary value and purchase frequency correspond to the size of the installed base whereas segment two (Potential) covers customers who purchase little spare parts in relation to the size of the installed base. Potential customers identified by the K-means algorithm are almost identical to the Potential customers identified by the Affinity propagation algorithm. Hence, both algorithms delivered quite similar results: they separated potential customers, i.e., customers that buy little spare parts compared to the size of the installed base.

The ScrewCo management representatives were involved in discussions regarding the results of the analyses. It is noteworthy that ScrewCo management perceived the identification of the potential customers particularly important:

"The analysis is good and it gives valuable information. Potential customers define a "focus list" that product managers can utilize when visiting markets (different local sales organizations). The list needs to be thoroughly examined. We have to find reasons for the list and find correct actions to decrease the number of potential customers. Small, Medium and Large customers seems logical, these customers do not require special treatment."

## 4.3 Market basket analysis

The data selected for market basket analysis covered sales of 177 different spare parts to selected customers from April 2011 to September 2016, in total 7,503 orders and tens of thousands of order lines. Orange data analysis tool (version 3.4.1) and Apriori algorithm was used for conducting the market basket analysis.

Table 3 highlights three interesting association rules found by the Apriori algorithm. The first rule means that 5 percent of orders (375 orders out of 7,503) included SP1 and SP2. In addition, if the order included SP 1 it was 91 % likely that the order included also SP2. The second rule means that 10 % of orders included SP3 and SP2, however the confidence was only 80 %. The last rule indicates that 1 percent of the orders (75) included SP4,SP5, SP1 and SP2. The confidence was 100 % meaning that all orders that included SP4, SP5 and SP1 included also SP2. In all these examples, Lift is over one meaning that there is a clear association between the given spare parts.

| *Antecedent* | | *Consequent* | *Support* | *Confidence* | *Lift* |
|---|---|---|---|---|---|
| SP1 | → | SP2 | 0.05 | 0.91 | 5.66 |
| SP3 | → | SP2 | 0.1 | 0.80 | 4.99 |
| SP4,SP5,SP1 | → | SP2 | 0.01 | 1.00 | 6.21 |

*Table 3. The market basket associations*

The management was positive about the found association rules and they confirmed the results:

"The results are very sensible. SP1 and SP3 are constantly wearing spare parts. Consequently, they are sold a lot and evenly. In addition, customers typically change SP2 at the same time with SP1 and SP3. This makes rule 1 and 2 very sensible. We should assure that this kind of analysis (market basket analysis) is part of the (upcoming) online shopping system. In our business, people typically rely on 'gurus' that recognize the equipment failure and the necessary spare parts of the device's operating noise. However, the analysis would in particular support 'non-guru' salespersons. It would be very interesting to see if different market areas produce different association rules"

## 4.4 Forecasting active customers

The data for forecasting active customers covered spare part sales to 485 customers between 2011 and 2016. Customer activity in 2016 was predicted with five variables: customer spare part volume in 2012-2015, number of days since last order at the beginning of 2016, Lumpy (1: if the customer had not made any order in any of the years 2012-2015, 0 otherwise), number of customer visits in 2012-2015, and the size of the customer's installed base at the end of 2015.

Scikit-learn machine learning library for Python and six different algorithms were used for classification: logistic regression (LR), linear discriminant analysis (LDA), k-nearest neighbor (KNN), classification and regression tree (CART), Naïve Bayes (NB) and support vector machine (SVM). Figure 3 shows results for different classification algorithms.

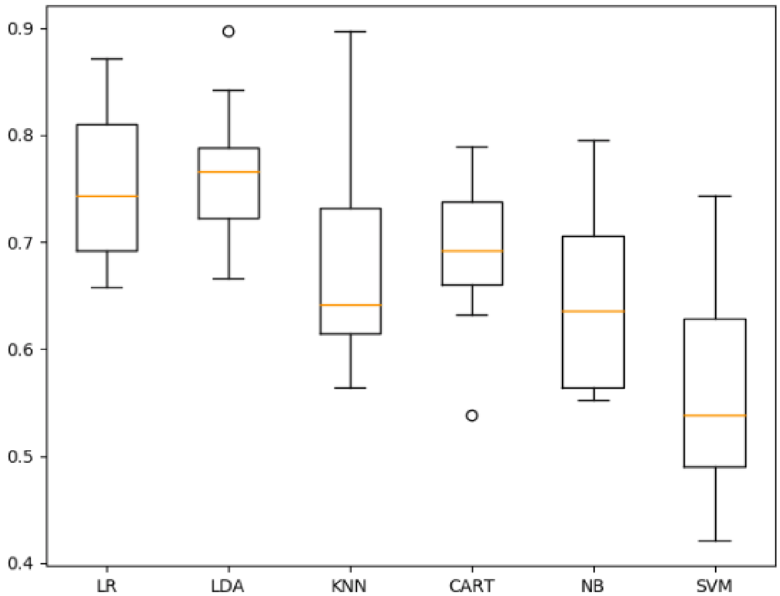


*Figure 3. Comparison of different classification algorithms*

The algorithm validation is based on 10-fold cross-validation. 10-fold cross validation divides data into ten subsamples. Every subsample is in turn retained as test set and the rest nine subsambles are used for training the algorithm causing each algorithm to run a total of ten times. The figure shows that algorithm results showed significant fluctuations. At the best, the prediction accuracy of the KNN algorithm was 90 % but the variance was high. The LR and LDA algorithms gave the most even results: the average prediction accuracy was around 75 %. Also, with respect to this analysis, management perceived that the results were interesting, but this analysis did not awaken immediate thoughts of further action.

## 4.5 Reflections upon the research diary

During the analyses, the researchers kept diary, in which they daily recorded and reflected upon the actions taken and the time elapsed. The diary was kept in order to find out the effort required to make the analyses from scratch. The idea was to imitate the situation where the company has the data needed for the analyses, but no experience of conducting the analyses. The researchers had previous programming experience. However, they had no earlier experience on ML and the used tools (Python and Orange).

The researchers got data in February 2017. Subsequently, researchers and the company members jointly chose the analyses to be carried out. The actual implementation started with the installation of the necessary tools in March 2017. The research project ended in June 2017 when the results were presented to the company members. Table 4 shows the research diary in more detail.

| *Timestamp* | *Action* | *Duration (hours)* |
|---|---|---|
| 21.3.2017 | Tools installation | 2 |

| | | |
|---|---|---|
| 22.3.2017 | Getting to know the tools and ML | 6 |
| 24.3.2017 | ML example | 3 |
| 27.3.2017 | Association rule: data manipulation | 3 |
| 28.3.2017 | Association rule: data manipulation and implementation | 3 |
| 28.3.2017 | Classification: Getting to know | 2 |
| 4.4.2017 | Classification: Data manipulation | 3 |
| 5.4.2017 | Classification: Data manipulation | 2 |
| 6.4.2017 | Classification: Implementation | 2 |
| 8.4.2017 | Classification: Implementation (Ready) | 3 |
| 13.4.2017 | Clustering: Data manipulation | 3 |
| 20.4.2017 | Clustering: Implementation | 6 |
| 27.4.2017 | Clustering: Implementation | 5 |
| 3.5.2017 | Clustering: Implementation (Ready) | 3 |
| May 2017 | Cleaning up the results | 8 |
| 2.6.2017 | Final presentation | 3 |
| | **Total** | 57 |

*Table 4. The research diary*

As the diary shows, the time spent on the analyses is rather limited, which (together with the positive feedback and implications) represent a positive sign for the feasibility and potential of the customer relationship analytics under examination in this study. However, the researchers had been collaborating with the company for many years and they knew the data and the company reporting practices in advance. In addition, the analyses made so far represent merely experiments and prototypes based on limited amount of data. Another challenge is to integrate the analyses into the company's CRM process, which requires much more work from several stakeholders.

# 5 Discussion and conclusion

The purpose of the paper is to evaluate the feasibility of B2C customer relationship analytics in the industrial B2B context. The paper responds to the literature gap on better understanding and managing the value potentials and dynamics within the B2B relationships (Anderson and Narus, 2003, Stormi *et al.*, 2018). The contribution of the paper is twofold; the article identifies analytics approaches with value potential for B2B decision-making and illustrates their value in use. The evaluated analytics, i.e., customer segmentation, market basket analysis and target customer selection, are based on different supervised and unsupervised ML algorithms. These forms of analytics are common in the B2C marketing where the number of customers is high (e.g. millions of customers), companies collect rich data from customers (especially in the eCommerce) and customer relationships are not often personal (Ngai *et al.*, 2009; Lin *et al.*, 2017). However, in the industrial B2B marketing the tested methods are not as common as the number of customers is relatively small (typically thousands) and marketing is more relationship than transaction based (Noori and Salimi, 2005; Martínez-López and Casillas, 2013).

In all, the potential of applying the advanced analytics lies in the possibilities to unveil the dynamics of the value of the B2B relationships and extended abilities to manage such dynamics. As the literature suggest (see e.g., Selos *et al.*, 2013), there are also many similarities between the B2B and B2C customers in terms of their people-driven decision-making, dynamics of loyalty and buying behavior. However, as illustrated by the case, in managing the B2B relationships, the companies may use new information sources about their installed base in use at the customers (Stormi *et al.*, 2018), and about the customers' business operations in a wider sense. As a context specific feature, in the industrial B2B context after sales (e.g. maintenance service and spare parts) are important. In addition, customer installed base is a prerequisite for the after sales. Consequently, installed base information can significantly enhance and support customer relationship analytics and help understanding the dynamics of business value related to those relationships.

The research also suggests that there are many practical implications in this area of research, essentially required in-depth understanding of the particular empirical context, in this paper industrial B2B after sales. The feasibility of the selected analytics, customer segmentation, market basket analysis and target customer selection, was evaluated with a single case study in industrial B2B after sales context from two perspectives: management judgement and implementation challenges. Overall, the managers considered the analyses interesting and useful. As the management said: “In future, our task is to find the right questions. I’m convinced that regardless of the question there is always a solution and method available to find the answer”

More precisely, as an example from the empirical context, customer segmentation revealed potential customers, i.e. customers that buy considerable less spare parts compared to their installed base. Hence, the use of customer installed base information increased the benefits of the segmentation analysis significantly. The management considered the market basket analysis as the most important. Market basket analysis identified spare parts that customers typically buy at the same time. In the case, the market basket analysis identified the most common spare part combinations as well as what spare parts customers usually buy in bundle. With this information, the sales can recommend spare parts to customers. Indeed, although the installed base, i.e., the equipment in use at the customers, determines overall demand for the spare parts, it is highly important information for the ScrewCo to understand what spare parts (among the hundreds of alternatives) are being ordered at the same time, by certain customers, customer segments and machinery in use in certain circumstances. Such information provides valuable input both for marketing management, and indirectly to the R&D efforts of ScrewCo. In long-term, in addition, ScrewCo is interested in selling spare parts also online where market basket analysis and recommender systems are commonly used.

As a practical limitation, in this paper, the researchers conducted the analyses within a short period as part of the wider research process. Due to earlier co-operation projects, the researchers had deep understanding about the company and the data and they also had programming experience. However, they had no experience on ML algorithms. To evaluate the implementation challenges, the researchers kept research diary. The most challenging part of the analyses was to edit the data in a suitable format and to get basic understanding about ML algorithms and the used tools. The analysis itself was relatively straightforward.

The research suggests that B2C customer relationship analytics are feasible also in the industrial B2B context and may serve different business purposes especially when the installed base information is combined with the analyses. Because B2B decision-making is also people driven, there is ample scope of future studies on employing B2C analytics in the B2B industrial context. Thus, this exploratory piece of research opens up many new research questions on the scope of the analyses and their practical implementation. *First*, the conducted analyses were based on static data that presented only small sample of the total customer data. The challenge is to get the analyses to work with the big amount of data in real life, in real time. *Second*, there are a numerous amount of other different analyses available, such as recommender systems, sequential analysis and anomaly/fraud detection. Recommender systems recommend products that meet customer unique profile and needs. Sequential analysis reveals the order of purchases. For example, customers typically first buy SP1, second SP5, and third SP6. Anomaly detection can reveal changes in customer buying behavior. *Third*, in the industrial B2B context installed base information offers far more opportunities than presented in the paper. Detailed information about customer’s installed base will help companies to recommend customers suitable products. In addition, pieces of equipment will gather more and more information about their operation. The data collected by pieces of equipment can move the customer relationship analytics to a completely new level.

## References

Aguinis, H., Forcum, L.E. and Joo, H. (2013). "Using market basket analysis in management research" *Journal of Management* 39 (7), 1799-1824.

Al-Htaybat, K., Al-Htaybat, K., von Alberti-Alhtaybat, L. and von Alberti-Alhtaybat, L. (2017). "Big Data and corporate reporting: impacts and paradoxes" *Accounting, Auditing & Accountability Journal* 30 (4), 850-873.

Anderson, J. C., and Narus, J. A. (2003). "Selectively pursuing more of your customer's business" *MIT Sloan Management Review* 44 (3), 42.

Appelbaum, D., Kogan, A., Vasarhelyi, M. and Yan, Z. (2017). "Impact of business analytics and enterprise systems on managerial accounting" *International Journal of Accounting Information Systems* 25, 29-44.

Berthon, P., Ewing, M., Pitt, L. and Naudé, P. (2003). "Understanding B2B and the Web: the acceleration of coordination and motivation" *Industrial Marketing Management* 32 (7), 553-561.

Bose, I. and Chen, X. (2009). "Quantitative models for direct marketing: A review from systems perspective" *European Journal of Operational Research* 195 (1), 1-16.

Bose, I. and Mahapatra, R.K. (2001). "Business data mining—a machine learning perspective" *Information & management* 39 (3), 211-225.

Chen, D., Sain, S.L. and Guo, K. (2012). "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining" *Journal of Database Marketing & Customer Strategy Management* 19 (3), 197-208.

Chen, I.J. and Popovich, K. (2003). "Understanding customer relationship management (CRM) People, process and technology" *Business process management journal* 9 (5), 672-688.

Chen, Y., Tang, K., Shen, R. and Hu, Y. (2005). "Market basket analysis in a multiple store environment" *Decision Support Systems* 40 (2), 339-354.

Choi, T.-., Chan, H.K. and Yue, X. (2017). "Recent Development in Big Data Analytics for Business Operations and Risk Management" *IEEE Transactions on Cybernetics* 47 (1), 81-92.

Chumpitaz Caceres, R. and Paparoidamis, N.G. (2007). "Service quality, relationship satisfaction, trust, commitment and business-to-business loyalty" *European journal of marketing* 41 (7/8), 836-867.

Cortez, R.M. and Johnston, W.J. (2017). "The future of B2B marketing theory: A historical and prospective analysis" *Industrial Marketing Management* 66, 90-102.

Dekker, R., Pinçe, Ç, Zuidwijk, R. and Jalil, M.N. (2013). "On the use of installed base information for spare parts logistics: A review of ideas and industry practice" *International Journal of Production Economics* 143 (2), 536-545.

Dursun, A. and Caber, M. (2016). "Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis" *Tourism Management Perspectives* 18, 153-160.

Fader, P.S., Hardie, B.G. and Lee, K.L. (2005). "RFM and CLV: Using iso-value curves for customer base analysis" *Journal of Marketing Research* 42 (4), 415-430.

Fisher, I.E., Garnsey, M.R. and Hughes, M.E. (2016). "Natural language processing in accounting, auditing and finance: a synthesis of the literature with a roadmap for future research" *Intelligent Systems in Accounting, Finance and Management* 23 (3), 157-214.

Gandomi, A. and Haider, M. (2015). "Beyond the hype: Big data concepts, methods, and analytics" *International Journal of Information Management* 35 (2), 137-144.

Gatzioura, A. and Sànchez-Marrè, M. (2015). "A case-based recommendation approach for market basket data" *IEEE Intelligent Systems* 30 (1), 20-27.

Geuens, S., Coussement, K. and De Bock, K.W. (2018). "A framework for configuring collaborative filtering-based recommendations derived from purchase data" *European Journal of Operational Research* 265 (1), 208-218.

Gordini, N. and Veglio, V. (2017). "Customers churn prediction and marketing retention strategies. an application of support vector machines based on the auc parameter-selection technique in b2b e-commerce industry" *Industrial Marketing Management* 62, 100-107.

Gordini, N. and Veglio, V. (2015). "Customer Relationship Management and Data Mining: A Classification Decision Tree to Predict Customer Purchasing Behavior in Global Market" In: *Marketing and Consumer Behavior: Concepts, Methodologies, Tools, and Applications*. Ed. by Information Resources Management Association (Ed.). IGI Global. Hershey, PA, USA. p. 789-828.

Güçdemir, H. and Selim, H. (2015). "Integrating multi-criteria decision making and clustering for business customer segmentation" *Industrial Management & Data Systems* 115 (6), 1022-1040.

Hosseini, S.M.S., Maleki, A. and Gholamian, M.R. (2010). "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty" *Expert Systems with Applications* 37 (7), 5259-5264.

Jiang, L., Cai, Z., Wang, D. and Zhang, H. (2012). "Improving Tree augmented Naive Bayes for class probability estimation" *Knowledge-Based Systems* 26, 239-245.

Kumar, V. (2010). "Customer Relationship Management" In: *Wiley International Encyclopedia of Marketing*. John Wiley & Sons, Ltd.

Lin, W., Ke, S. and Tsai, C. (2017). "Top 10 data mining techniques in business applications: a brief survey" *Kybernetes* 46 (7), 1158-1170.

Lindholm, A., Laine, T. J., and Suomala, P. (2017). ”The potential of management accounting and control in global operations: Profitability-driven service business development” *Journal of Service Theory and Practice* 27 (2), 496-514.

Martínez-López, F.J. and Casillas, J. (2013). "Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights" *Industrial Marketing Management* 42 (4), 489-495.

Nairn, A. and Bottomley, P. (2003). "Something approaching science? Cluster analysis procedures in the CRM era" *International Journal of Market Research* 45 (2), 241-264.

Ngai, E.W., Xiu, L. and Chau, D.C. (2009). "Application of data mining techniques in customer relationship management: A literature review and classification" *Expert Systems with Applications* 36 (2), 2592-2602.

Noori, B. and Salimi, M.H. (2005). "A decision-support system for business-to-business marketing" *Journal of Business and Industrial Marketing* 20 (4-5), 226-236.

Oliva, R. and Kallenberg, R. (2003). "Managing the transition from products to services" *International journal of service industry management* 14 (2), 160-172.

Olson, D.L. and Chae, B.K. (2012). "Direct marketing decision support through predictive customer response modeling" *Decision Support Systems* 54 (1), 443-451.

Pilerood, A.E. and Gholamian, M.R. (2016). "An efficient customer classification framework to identify target customers by Bayesian networks in automotive industry" *International Journal of Data Analysis Techniques and Strategies* 8 (4), 348-356.

Qiu, J., Lin, Z. and Li, Y. (2015). "Predicting customer purchase behavior in the e-commerce context" *Electronic commerce research* 15 (4), 427-452.

Ramya, V. and Ramakrishnan, M. (2016). "Mining Association Rules from Large Volumes of Data: A Survey" *International Journal of Control Theory and Applications* 9 (34), 459-465.

Reimer, K. and Becker, J.U. (2015). "What customer information should companies use for customer relationship management? Practical insights from empirical research" *Management Review Quarterly* 65 (3), 149-182.

Russell, G.J. and Petersen, A. (2000). "Analysis of cross category dependence in market basket selection" *Journal of Retailing* 76 (3), 367-392.

Sarvari, P.A., Ustundag, A. and Takci, H. (2016). "Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis" *Kybernetes* 45 (7), 1129-1157.

Schläfke, M., Silvi, R. and Möller, K. (2012). "A framework for business analytics in performance management" *International Journal of Productivity and Performance Management* 62 (1), 110-122.

Selos, E., Laine, T., Roos, I., Suomala, P., and Pitkänen, L. (2013). ”Applying SPAT for understanding B-to-B supplier switching processes” *Managing Service Quality: An International Journal* 23 (4), 321-340.

Stormi, K., Laine, T., Suomala, P. and Elomaa, T. (2018). "Forecasting sales in industrial services: modeling business potential with installed base information " *Journal of Service Management* 29 (2), 178-205.

Syed, A., Gillela, K. and Venugopal, C. (2013). "The future revolution on big data" *International Journal of Advanced Research in Computer and Communication Engineering* 2 (6), 2446-2451.

Trkman, P., McCormack, K., De Oliveira, Marcos Paulo Valadares and Ladeira, M.B. (2010). "The impact of business analytics on supply chain performance" *Decision Support Systems* 49 (3), 318-327.

Warren Jr, J.D., Moffitt, K.C. and Byrnes, P. (2015). "How Big Data will change accounting" *Accounting Horizons* 29 (2), 397-407.

Weinstein, A. (2014). "Segmenting B2B technology markets via psychographics: an exploratory study" *Journal of Strategic Marketing* 22 (3), 257-267.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B. and Philip, S.Y. (2008). "Top 10 algorithms in data mining" *Knowledge and information systems* 14 (1), 1-37.

Wübben, M. and Wangenheim, F.v. (2008). "Instant customer base analysis: Managerial heuristics often “get it right”" *Journal of Marketing* 72 (3), 82-93.