WHITEPAPER

Big data in banking

for marketers

How to derive value from big data

BANK 2020

EVRY

FOREWORD

In Marketing & Sales the main strategic goals are to acquire new customers, develop as well as to retain existing ones. Financial service organizations generate huge amounts of data such as purchase history, profile data, browsing history or social media data every single day. Used in a wise way, marketers can employ data analytics to reach their main marketing objectives more efficiently.

During the last years Big Data has become the buzzword across various industries, but it is difficult to know exactly what Big Data can do to improve business value and which Big Data applications marketers should consider to invest both their time and money in.

Our goal in this paper is to show a comprehensive list of data-driven use cases and their value, which are deployed by successful marketing teams today. Each use case is chosen for its relevancy to the banking sector and is backed by case studies from real organizations.

We hope you enjoy reading it as much as we enjoyed making it.

Big Data Research @ EVRY

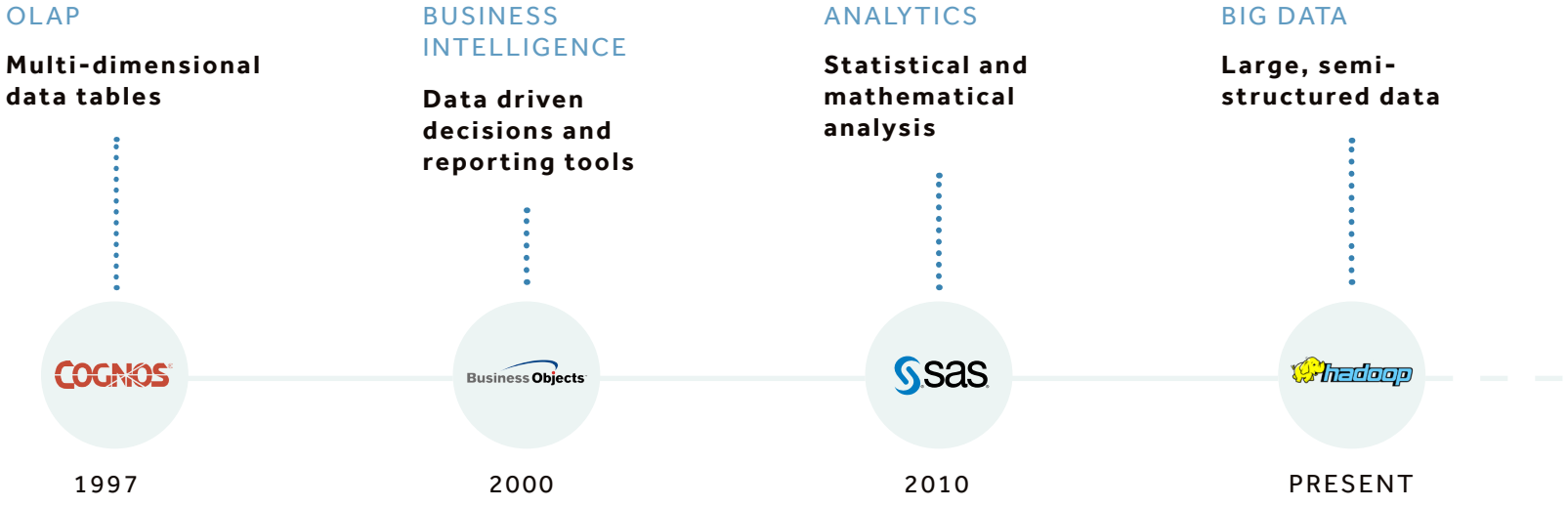
THE DEFINITION
OF BIG DATA

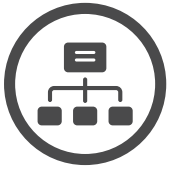
One of the key problems initially faced with any Big Data research is that the term itself is highly ambiguous and a quick Google search reveals millions of pages each with their own spin on the definition of Big Data. This can lead to misunderstandings when talking about the topic. To avoid any confusion we outline our own Big Data definition right here:

Big data was born out of the necessity of data sets growing so large and complex that traditional tools are no longer sufficient to process this data. By aggregating large amounts of data from many different sources makes big data very powerful for business decision-making, revealing insights and behaviours faster and better than otherwise possible with traditional BI.

THE TRENDS OF DATA ANALYTICS

Big Data has been making all the headlines over the last few years, but it is just one data trend in a long line that appear every decade. The difference between these data analytic developments is blurred at best, but the underlying technology does appear to give some distinction between the old and the new trends. The technology behind Big Data is undeniably Hadoop, a software ecosystem designed to allow the query and statistical analysis of large and semi-structured data. Hadoop's ability and flexibility to handle increasingly complex data has unlocked new opportunities for extracting value and business insights from potentially massive amounts of organizational internal data. Big Data has also allowed the possibility to enrich this internal data with equally vast amounts of semi-structured external data from public sources and social media, maximising data value potential even further. Combining and processing all this internal and external data was simply not possible using traditional management and analytical tools before the age of Big Data.



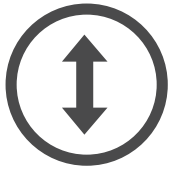


DATA

The data can range from structured to completely unstructured. Often it has a lot of metadata attached, which can be as equally useful as the data itself. The scope of data sources can be especially large, across enterprise databases and internet channels.

TECHNOLOGY

Technologies such as the Hadoop ecosystem have been designed from the ground up to be able to handle large and complex data, which are the key features of Big Data. What makes Hadoop particularly powerful is its ability to linearly scale with increasing data complexity, making it an almost invaluable tool in any Big Data application.



SIZE

Big data invariably contains large amounts of data, but how much is large? There are no straight answers, but if the amount of data is breaking traditional systems and Hadoop is needed to be used, then it is Big Data.

THE ATTRIBUTES OF BIG DATA

SOME WORDS ON HOW TO USE THIS WHITEPAPER

The following pages will outline different Big Data use cases relevant for marketing managers in the banking industry. This is designed in a way that areas of interest can be found easily without reading from the first until the last page. Every use case will first summarize the top uses of the Big Data application before each one is explained in more detail. In this booklet we try and focus on Big Data company examples, however traditional data analytic cases have been included because of their relevance or because they offer a high data to value potential.
The whitepaper is supposed to be a quick and easy read, therefore icons are used for the different algorithms explained.

Look out for these symbols for the algorithms:

DECISION TREES & RANDOM FORESTS

PAGE 55

CLUSTERING

PAGE 56

TEXT ANALYTICS

PAGE 57

NEURAL NETWORKS

PAGE 58

LINK ANALYSIS

PAGE 59

SURVIVAL ANALYSIS

PAGE 60

CONTENTS

5 USE CASES TO

ACQUIRE CUSTOMERS

DEVELOP CUSTOMERS

RETAIN CUSTOMERS

1 SENTIMENT ANALYTICS

» Monitor what customers say to increase marketing success
» Identify key customers to boost word-of-mouth marketing
» Examine customer feedback to improve products and services

PAGE 7

2 CUSTOMER 360

» Identify the customer profile
» Understanding the product engagement of the customer
» Detect when a customer is about to leave

PAGE 15

1

USE CASE

SENTIMENT ANALYTICS

Social media has opened new avenues and opportunities for organizations to connect with their customers, but the sheer volume of communications about brands, products and services; discussed, shared, criticised or liked on different social platforms can be overwhelming. Sentiment analytics helps to rapidly read all this data, providing an executive summary of what people like and don't like about a company brand or products. The reasons behind the sentiment can then be easily extracted, providing valuable business insights.

TOP USES
FOR SENTIMENT ANALYTICS

1.1

MONITOR WHAT CUSTOMERS SAY TO INCREASE MARKETING SUCCESS

Meaningful data includes opinions, feelings and attitudes about a brand, topic or keyword, which are shared freely in the world wide web. Knowing the opinion and attitude of the customers offers profound knowledge in order to adjust marketing tactics correctly.

PAGE 9

1.2

IDENTIFY KEY CUSTOMERS TO BOOST WORD-OF-MOUTH MARKETING

The right sentiment analysis tool can also identify the most influential customers regarding company brands or products. It enables to engage with the right people who will spread the word and has much influence on a social platform. Exactly those key customers are critical in order to fulfil the goals for a successful acquisition strategy.

PAGE 11

1.3

EXAMINE CUSTOMER FEEDBACK TO IMPROVE PRODUCTS AND SERVICES

Many consumers freely give feedback and product suggestions on social media websites. Big Data technologies can be used to identify those valuable consumer insights to improve products and services much faster than with traditional surveys, which only portrays a small sample group at one specific moment in time.

PAGE 13

MONITOR WHAT CUSTOMERS SAY TO INCREASE MARKETING SUCCESS

Sentiment analytics or opinion mining offers a deeper understanding of the crowd. It gives a focus group of millions of customers that can be tracked, revealing insights in how they think about a certain topic.

With the help of sentiment analysis it is possible to understand the sentiment of customers and know what they say about new products, services, branding or commercials. Sentiment analysis uses natural language processing and text mining techniques to read vast amounts of information that would be impossible for even a large team to give an executive summary and insights hidden amongst the data.

BBVA

BBVA offers a wide range of finanical and non financial service products to its approximately 47 million customers across 30 countries. BBVA uses social media to determine how customers feel about their brand and their products.

This information is used to adjust marketing tactics in order to increase positive feedback and customer satisfaction. BBVA can then report the success of its media messaging and the causes for positive and negative sentiment can be identified.

KIA

The car manufacturer Kia Motors is always eager to find out what customers think and say about their cars. The company uses sentiment analytics tools to detect what is said about the brand and products on various blogs, Twitter and Facebook.

For example: Big Data technology has enabled the marketing team to examine the power of an advertisement spot during the Super Bowl game, based on the reactions on social media platforms.

NEDBANK

Nedbank Ltd is a large bank in South Africa that realizes great advantages by using social media analytics. Analysing various social media platforms in almost real-time provides Nedbank's marketing department information about the marketing campaign, customer preferences and complaints.

This technology implementation has decreased social media monitoring costs significantly while enhancing marketing success.

HOW

Sentiment analytics is simply the study of customer's sentiment towards any object of interest. Sentiment analytics is typically done over a period of time to measure the success of a brand campaign or how rival brands are perceived compared to another company. Most data available to an organization to measure customer's sentiment will be available in text format such as from social media. As it would be humanly impossible or highly impractical to read all the comments in the social sphere on review sites and blogs to assess customer sentiment, text analytics algorithms, such as naïve Bayes, are ideal for this type of problem. Naïve Bayes can analyse document on its level negative and positive sentiment and produce a score typically in the ranges from highly negative (-1) to neutral (0) to highly positive (1). This scoring allows an organization to view to overall sentiment without having to read all the comments posted about their products or brand.

ALGORITHMS*

*See page 52 for Appendix

DATA

An easy source of customer sentiment is from the social sphere, including social networks, blogs and review sites. This data is naturally unstructured and dynamic as new text is generated continuously. This data is then suited to measuring sentiment over time such as before and after an organizations branding efforts. Internal data gathered from past consumer surveys and call logs may also provide a good source to measure the customer sentiment towards particular products.

IDENTIFY KEY CUSTOMERS
TO BOOST WORD-OF-MOUTH MARKETING

By targeting customers with high influence, product awareness can be rapidly increased, since word-of mouth recommendation is still the most trusted form of advertisement.

A great product should sell itself in an ideal world, but even in our world this still holds true to a certain extent as word-of-mouth is a very powerful tool in the marketing arsenal. It works great for an established product with an already entrenched fan base, but for new products this can pose a problem. New products need a method to kick-start the product awareness amongst influential customers who are likely to build up a momentum of interest amongst the potential customer base.

The Ford Fiesta was a big hit in Europe in 2010, but had yet to break the American market. To help maximise the launch of the Fiesta in the USA, Ford identified 100 of the most influential bloggers and recruited them into its "Fiesta Movement" campaign, where they were each given a Fiesta and missions such as driving to music festivals. This approach generated a lot of buzz around the car at a much lower cost than a traditional television-centred method.

Starbucks is something of a pioneer when considering the importance of influential customers. Almost all Starbucks advertising is done by word-of-mouth and even new product ideas are generated by the customer base using the mystarbucksideas.com portal.

T-Mobile understands that customers can sometimes behave like dominoes. It only takes one highly influential customer to switch to another network provider before large numbers of customers follow in their wake. To try and find out who the valuable high influential customers are, T-Mobile uses a mixture of customer influence scoring and customer value KPI's to build an influence profile of each customer. T- Mobile can then target these customers for special attention and take appropriate actions.

HOW

If highly influential customers are already known, such as those that have been flagged for a high number of referrals then decision trees can be used to determine the most important attributes of an influential customer. If this information is unknown, the influence of customers in the social sphere can be determined by a combination of their reach and level of interaction they have with other people.

Link analysis and graph theory can provide an accurate measure of reach by finding which customers act as nodes in the wider social network. Decision trees can then be used to evaluate the level of interaction this customer has with others. Often each customer is then provided a leadership score, with high scoring leaders offered to try out new products, services or features to spread the product awareness or customers may also be flagged for special attention.

ALGORITHMS*

*See page 52 for Appendix

DATA

Internal customer profile may reveal which customers or clients give a high number of referrals. This database may also contain the attributes that can be used to determine if the customer is influential, such as demographic information. Link analysis strongly depends on highly unstructured social network data and data from third party blogs and review sites.

EXAMINE CUSTOMER FEEDBACK

TO IMPROVE PRODUCTS AND SERVICES

Many products and services can be easily improved, if a direct communication with customers exists. Social media platforms are great sources of communications to discover improvement opportunities, banks just need to filter these insights out from the torrent of information.

Traditional market research tools like customer surveys or focus groups are not only time and cost consuming, they are also not accurate since the sample group is rather limited. Sentiment analysis tools make use of huge amounts of communication data from social media networks and logs, helping to improve products and services promptly.

BARCLAYS

Barclays was able to derive actionable insights from real-time social media analytics after they launched their new mobile banking app.

The app did not allow young consumers under 18 to transfer or receive money. This created negative comments from teenagers as well as from their parents as they could not transfer any money to their kids. After the data revealed this problem, Barclays improved their app promptly, adding access for 16 and 17 year-olds.

DELL

The computer manufacturer Dell improves its products based on sentiment analytics, which enables the company to listen carefully to its customers on the world wide web.

Negative feedback about overheating on a newly released laptop helped them, for example, to overcome the problem quickly and to create a better product for future customers.

intuit

Intuit, an American company that provides a tax service software called Turbo Tax uses sentiment analytics to improve its product continuously. "Tax season only lasts a hundred days, and we measure [customer attitudes to TurboTax] every day, using what we hear to drive improvements all the way to April 15. We don't have time to go through all the customer comments manually, so we need a sentiment analysis tool that we can rely on and trust."

HOW

Nearly all sources of customer sentiment will be available in a text based format, either from social media networks or comments written by staff in call centres and brick and mortar stores. The Naïve Bayes algorithm is an extremely popular method for classifying text and documents and is well suited to finding positive and negative sentiments amongst a customer base. Once these sentiments are found, they can be filtered based on product or service.

The top 5-10 words that lead to negative or positive sentiment can then be extracted from the Naïve Bayes algorithm for each product and service. These key words may provide insight on how the product can be improved. For example the word time might be flagged for a service, which can then form the basis of an investigation of why time is causing this service to underperform.

ALGORITHMS*

*See page 52 for Appendix

DATA

The social sphere including the own Facebook homepage and twitter can form the basis of data source for customer sentiment. Twitter is particularly useful as its openness allows to search all tweets across the entire network rather than those just on the homepage.

2

USE CASE
CUSTOMER 360

Understanding the customer as a whole is important to stay ahead of competitors.

There are several important aspects to consider when developing a 360-degree customer-view. The past and immediate customer behaviour is important to predict future customer trends and what their most likely next action will be. The customer's transactions and travel habits are also important to build a lifestyle profile and discover new insights. These are just some customer attributes that are used to build a complete and holistic customer picture.

TOP USES
FOR CUSTOMER 360

2.1

IDENTIFY THE CUSTOMER PROFILE

Knowing the customer profiles means having a deep understanding about them, which can be used to drive actionable insights. This can lead to improved marketing campaigns, targeted sales and better customer service. A clearer view about the customer profile enables companies for example to send out triggered messaging, which is a good way to reinforce the brand and target customers.

PAGE 17

2.2

UNDERSTANDING THE PRODUCT ENGAGEMENT OF THE CUSTOMER

Understanding how consumers are using a specific product and then making decisions accordingly makes a big difference. With the help of Big Data analytics companies can find out how engaged a consumer is with a product. This can help to send the correct marketing message and product when the customer needs or wants it most.

PAGE 19

2.3

DETECT WHEN A CUSTOMER IS ABOUT TO LEAVE

Analysing customer behaviour is not just reviewing past historical purchases, but a tool to forecast future actions and trends of customers. By predicting customer behaviour, insights can be revealed to stop churn before it is too late.

PAGE 21

IDENTIFY THE CUSTOMER PROFILE

Knowing the individual profile of the customer allows to send out personalized marketing messages. A small touch of personalization can improve customer engagement, security and loyalty.

Personalization within marketing messaging can be used to achieve a number of goals. This might be as simple as being more friendly by adding the customer's name to the start of the message or by adding customer details only the organization would know to prevent fraudulent marketing messages. More ambitious personalization would be to add personalized content within the message that would directly interest the customer increasing its engagement.

HDFC BANK
We understand your world

HDFC Bank uses customer lifecycle events to boost credit card activations. This is achieved by targeting promotions with personalized messages to each of the lifecycle segments that HDFC had identified.

The result of this is a significant increase in the number of credit card activations and a reduction in cost per acquisition of each customer.

hmv

HMV, a British entertainment retailing company realized that sending the same campaign message to all its customers is not appropriate anymore, as people start treating emails as spam and do not open them. The company uses a recommendation system, which analyses customer click streams and which products fits the customer's preferences. HMV sends out personalized recommendations, which increased the emails opening by over 70% on mobile phones, and PC mails by 50 %.

OCBC Bank

OCBC is currently the largest local bank by market capitalisation in Singapore and operates in 15 countries world wide. By responding to the customer actions, personal lifetime events and demographic profiles OCBC Bank is able to achieve higher customer engagement and to increase customer satisfaction by 20% compared to a control group.

HOW

Classification algorithms such as neural networks or decision trees can be used to help determine what content the customer may or may not be interested in. These may be lifestyle choices like if the customer is a frequent traveller or life stage events such as is the customer a student. This enables the campaign manager to selectively add information from the organizational blogs or products that may be of interest to the customer. Recommendation engines using algorithms such as linear regression and nearest neighbours are used to compare customers to their peers in a similar fashion to the HMV example. This can also be combined with logistic regression to measure the probability that a customer will click an advertisement in the context of the product being advertised. Together these algorithms provide powerful insights into how a customer will respond to a marketing message.

ALGORITHMS*

*See page 52 for Appendix

DATA

Customer profile is primarily used to add personalization to marketing messages. This includes the customers demographics such as their post code and their name. Transactional data may also reveal the customers interests, which can be used to help add relevant information to the marketing messaging.

UNDERSTANDING THE PRODUCT ENGAGEMENT OF THE CUSTOMER

How the customer uses or doesn't use a product can be a sign on how to adjust the marketing message.

Understanding where the customer is in the product lifecycle can have significant impact on the type of marketing communications that is used. For example customers in the acquisition stage are more susceptible to marketing and product research messaging than an ingrained customer. A customer about to churn would make a better candidate for retention offers than a loyal customer. Big data can help to identify in which stage of the product lifecycle a customer is in and helps to adapt the message accordingly.

FIRST TENNESSEE
powering your dreams™

The First Tennessee Bank uses product lifecycle stages such as loyal customers, high value customers, apprehensive customers and wavering customers to name just a few key segments. For example, re-engagement messaging may be useful for past high value customers who haven't been in contact with the bank for a while. Loyal and engaged customers may prefer more VIP offers.

Bank Austria
UniCredit Group

Also Bank Austria makes use of understanding the product lifecycle to retain their customers. When a customer shows specific behaviour connected with cancelation of a product, the banking staff detects this and takes according actions for a renewal.

Jeanswest

The retail company Jeanswest knows their customer engagement well and sends out appropriate offers. For example frequent shoppers receive a thank you offer to show them they are valued, new customers get a welcome offer and irregular shoppers are encouraged to shop again at Jeanswest.

HOW

The product lifecycle can be broken into three broad categories, acquisition, activation and relationship management. Analysing how often the customer interacts with the company can be a good indicator of which group they belong to. This is simply a classification problem, which can be solved either using neural networks or decision trees. The customers interaction across multiple channels would also need to be taken into consideration as they are likely to move between channels as their lifecycle develops. Nearest neighbour and linear regression can also be powerful tools on classifying the customers to product lifecycle stage by comparing them to their peers.

ALGORITHMS*

*See page 52 for Appendix

DATA

To accurately assign a product lifecycle stage to a customer, profile data and transaction data is needed to determine how they use the products. Metrics such as customer engagement can be gained from this data. Recent clickstream URL referrals and cookie data gives information if the customer is in the research stage of the product cycle.

DETECT

WHEN A CUSTOMER IS ABOUT TO LEAVE

Predicting which customers will churn should form a crucial part of an organizations customer orientated strategy as the cost of acquiring new customers is far greater than the cost of keeping the existing customers.

Retaining consumers is easier when an organization understands the customer in a holistic way. Indicators that might point to attrition can be cancellation of automatic payments, customer complaints on customer calls or on social media. By analysing the different indicators, potential customer churn can be identified before it is too late.

American Express is using sophisticated predictive models to analyse historical transactions and 115 variables to forecast customer churn. The company believes it can now accurately identify 24% of accounts within its Australian branch that will close within the next four months and can take according steps to prevent them from churning.

Tatra Bank, Slovakia's first private bank has nearly reached its target of decreasing customer churn from credit card holders by 30 percent with the use of predictive models. The bank segments their customers and has selected highly personalized retention campaigns for their customer groups.

Customer data of T-Mobile USA includes the time and lengths of call, internet usage or peak times for direct messaging. T-Mobile USA takes advantage of this data to prevent customer churn. An example of this is billing analysis, where the product usage is calculated. If the frequency of calls to contacts who are using a new providers are increasing this could imply that friends or family are switching providers, and the customer might possibly do so as well. By identifying these customers T- Mobile USA achieved to decrease their churn rate by 50% in just one quarter.

HOW

Customer churn can be modelled in a number of ways, but a popular method is to use a random forest or decision tree approach. The output of such algorithms is typically a churn score or a probability that a customer is likely to churn. The advantage of these algorithms is that they can handle a wide array of fields and data types that all might contribute to the churn of the customer. These algorithms also intuitively show which fields are the most important when predicting churn. This can be invaluable information when asking why customers are churning and not just who will churn. Survival analysis is another popular approach to predicting when customers will leave and is a powerful tool in predicting increase and decreasing hazards in customer survival. The advantage of using survival analysis is that it is easy to compare different customer segments across a time series. This may reveal that churn is different amongst older customers and younger customers and so a different model should be used to model churn between these two customer segments.

ALGORITHMS*

*See page 52 for Appendix

DATA

Customer profile and transaction data can be used to analyse how the customer is using the product, which in turn can be used as behavioural indicators of potential churn. Social data from sentiment analysis may also indicate if particular segments are unhappy with an organizations service and are likely to churn.

3

USE CASE

CUSTOMER SEGMENTATION

Segmentation is simply dividing the customers into natural groupings that share similar characteristics or behaviours.
Understanding these groups is necessary to find needs and wants, that forms the basis for a sales & marketing strategy.
The concept of segmentation is not new and frequently used in companies, but Big Data facilitates to create sharper segments faster.
It helps to see the existing customer base in new ways, which creates unique business opportunities.

TOP USES
FOR CUSTOMER SEGMENTATION

3.1

DESIGN TARGETED MARKETING PROGRAMS

Big Data segmentation allows to see how customers are really using products and what issues they care about most. This enables companies to discover segments that have traditionally been underserved. Highly optimized marketing messages for each of the groups can then be developed, creating greater resonance with customers.

PAGE 25

3.2

CREATING LOYALTY PROGRAMS BASED ON CARD USAGE HABITS

Accurately defining segments enables banks to provide highly personalized cash back offers with vertical partners such as food stores, retail or travel companies. This can significantly increase card loyalty and card usage.

PAGE 27

3.3

OPTIMIZE PRICING STRATEGY

Knowing how much a segment is willing to pay for a product or service is a key business strategy for many companies. Big Data segmentation allows to find more of these groups and their price willingness, creating a sliding scale optimized pricing strategy for the customer base.

PAGE 29

3.4

BUILD RELATIONSHIPS WITH VALUABLE CUSTOMERS

With the help of Big Data, profitable market groups can be identified and given preferential treatment to strengthen customer satisfaction. Segmentation also reveals the attributes of a profitable person, enabling banks to target high profit potential customers.

PAGE 31

DESIGN TARGETED MARKETING PROGRAMS

Targeting the right market groups with the right offers can significantly increase marketing effectiveness.

Big Data analytics is able to consider aspects beyond traditional segmentation of customer age and martial status and predict the customers groups based on lifestyle, life stage and special events. This allows for new and more representative cohorts of a customer base that reveals the needs and wants with far greater accuracy. The information provides a marketing team the opportunity to design more personalized marketing programs to an organization's segment that may have historically been underserved or ignored. Once these groups have been discovered, new customers can be classified into an appropriate section in milliseconds opening up new opportunities in real time advertising.

Bank of America

By segmenting Bank of America is able to remove its assumptions about its customers. This led to a change in its marketing message from "Use the value of your home to send your kids to school" to "Use the value of your home to do what you always wanted to do" and increased conversion rates tenfold.

First Tennessee

First Tennessee Bank combines their knowledge on segments with real P&L data to improve its marketing effectiveness, concentrating on programs that deliver the maximum ROI. Better targeted marketing programs has led to a 3,1% higher response rate and marketing costs could be decreased by almost 20%.

BBVA

BBVA makes use of understanding the customer lifecycle to target the right offer to the right customer. The bank groups consumers according to life stages, such as singles, independent professionals, young families and retirees.

This information is than used to serve this segment accordingly with the appropriate marketing message and products.

HOW

The problem of targeted advertising campaigns is finding out what the customer really cares about. This is not a trivial problem, as it requires the use of clustering algorithms to find the meaningful patterns within the customer data and then human intuition to link these patterns to real customer behaviours. For example, in the Bank of America case, data was taken from the customers in the target demographic who had home equity loans. From this data researchers found a large cluster of customers that had both home equity loans and their own business. It took further research to identify this pattern was caused by customers using home equity to build their businesses. In this example the K- means clustering algorithm was used, which is likely the most popular algorithm for this type of problem.

ALGORITHMS*

*See page 52 for Appendix

DATA

All customer data can be used for cluster analysis, but better results if the data is truncated to a particular customer demographic or product users. An example would be to investigate students that use savings accounts. For that the retiree demographic would not add any meaningful insights.

CREATING LOYALTY PROGRAMS

BASED ON USAGE HABITS

Customers can be readily grouped according to their card usage habits, which can be used to offer them more personalized loyalty offers.

Segmentation based on general transaction data such as purchase frequency and purchasing habits can reveal lifestyle segments such as fashion conscious-, family orientated-, or travel oriented groups.
This allows for a marketing team to design loyalty programs with vertical partners around what their customers really want. For example, family orientated customers might appreciate rebates at supermarkets and gas stations.

citibank

Singapore Citibank offers customer discounts at retailers and restaurants based on the customer transactional patterns. By offering this service, Singapore Citibank has a significant increase in its card usage loyalty, retention and overall improvement of customer satisfaction.

RBC

The Royal Bank of Canada (RBC) designs its loyalty programs based on its customer's card usage habits. This allows the bank to offer their customer's tailored products, including gifts for pets like carriers, dog feeders and doggie loungers.

Bank of America

Bank of America is employing targeted marketing programs to increase the card usage of its customers through its Bankamerideals loyalty program. The Bankamerideals loyalty program includes tailored customer-centric rewards and charity-of-choice donation.

HOW

To design effective loyalty programs banks need to know what their customers consider important. A clustering algorithm, such as the popular K-means clustering, helps solve this problem by uncovering meaningful patterns amongst the customers profile and transactional data. Examples of behavioural clusters using a banks transaction and profile data include telco users, family man, fashion ladies, etc. Although the algorithm found these clusters based on meaningful patterns, it took human intelligence to link these clusters to the underlying customers behaviour. These clusters can then form the segments of your loyalty based programs, maximising the effectiveness of the loyalty program with each segment.

ALGORITHMS*

*See page 52 for Appendix

DATA

Transaction data is primarily used for this use case, but customer profile data can help narrow the aim is to design a loyalty program for a particular demographic.

OPTIMIZE
PRICING-STRATEGY

Knowing what customers are willing to pay for a product or service can be a complex task, but the rewards of accurately predicting their maximum price point can significantly increase ROI.

Taking advantage of data-driven pricing strategies needs technology that can capture and manage large customer data sets, breaking it down in different segments based on the customer's willingness to spend on a product or service. Banks can identify the right pricing levels for these customers by identifying the pricing elasticity for each segment. Certain segments might receive automatic exception pricing, or will get offered an alternative product at a more competitive price point. These strategies are critical in order to maximize profitability and ROI on the existing portfolio.

RYANAIR THE LOW FARES AIRLINE

Customer segmentation for an optimized pricing strategy is widely used in the tourism industry for example by airline companies or tour operators. Airlines, such as Ryanair use algorithms to segment customers to optimize their ticket price, maximising their ROI without being perceived as being too expensive. This enables Ryanair to stay competitive whilst maintaining their low cost ticket mantra.

الإتحاد ETIHAD AIRWAYS

Ethiad Airways, which carries over 10 millions passengers annually is using Big Data to optimize their pricing strategy. They have achieved this by making use of their frequent flying programme to develop strategic pricing. By examining the frequent flying behaviour, the company can improve pricing by tracking upgrade frequency and other transactions.

FIFTH THIRD BANK

Fifth Third Bank uses analytics-based product pricing engine to help acquire new customers. Using data analytics the bank can run scenarios on how various price points will influence its customer acquisition and deposit levels. For example, the bank can make price predictions when interest rates will rise in the future and make scenarios where it wants to be with rates in the market to be aggressive in attracting customers.

HOW

Segmenting customers based on price discrimination is initially a clustering algorithm problem followed by a classification problem. Clustering can first identify the spending habit of certain groups, lifestyles or demographics of customers to calculate their willingness to spend on certain products. Such clusters calculated from banks data include, heavy spenders, one time buyers, instalment buyers etc. Each of these clusters can then be assigned an optimal price to maximise ROI. Once these clusters are identified and assigned a price, a classification algorithm such as a decision tree is then needed to assign customer prospects to one of the clusters. Using a classification algorithm post clustering prevents the need to continually recalculate the clusters with every new customer prospect, maximising the speed of the algorithm.

ALGORITHMS*

*See page 52 for Appendix

DATA

Transaction history and customer's profile are needed to view the customer's behaviour and what their spending potential is. Detailed customer profile can allow price clustering based on demographics and lifestyle behaviours.

BUILD RELATIONSHIPS
WITH VALUABLE CUSTOMERS

In todays competitive market, banks need to find out which segments are strategically valuable. Then extra care needs to be taken to strengthen the customer relationship with these valauble segments.

Segmentation can be used to identify loyal and profitable groups but beyond this, it can also find which segment is likely to evolve to an important one in the near future. Armed with this knowledge companies can strive to increase the market share of these more profitable and high potential customer groups while decreasing investment into low profit segments.

CAESARS ENTERTAINMENT

Caesars Entertainment, owner and developer of Caesars Entertainment casinos, takes advantage of Big Data analytics to improve its customer relationship with their most loyal customers. The company is using customer data from its loyalty program, web clickstreams and from real-time play in slot machines to identify the most profitable and loyal customers, making sure they are satisfied at all times.

BARCLAYS

Barclays is making use of segmentation by targeting students with personalized advertisement and attractive offers. The objective of reaching out to this market segment is straightforward. Students most likely pick a bank for the first time and Barclays tries to keep these customers in the long term when they will become profitable customers.

ZIONS BANK

Zions Bank employs Big Data to derive insights about the growing Hispanic community. According to Bank Technology News, data analytics helps "the bank identify Hispanic consumers within its geographic footprint and see their education levels, home ownership, affluence and other factors. The bank then breaks the segment down into groups and identifies financial needs for each." Based on these segmentation insights the bank could strengthen the customer relationship and has grown its Hispanic customer base 10-15% for the past few years.

HOW

Dividing customers based on their profitability is a popular application in customer segmentation. Like most segmentation-based tactics, a clustering algorithm such a K-Means can be used to find natural groupings of valuable customers and which traits lead to these customers being significant. A valuable customer can be defined by a KPI such as a standard profitability measure, which takes a snapshot of a customer at a single point in time to define its profitability. The customer lifetime value also makes a good profitability measure that indicates how the prosperity of a customer might increase or decrease over time. To calculate these KPIs the customers churn rate, retention cost, profit by unit sale, discount rate and historical value are needed.

ALGORITHMS*

*See page 52 for Appendix

DATA

Some key data is needed to measure customer profitability. This includes demographics, the type frequency and occasions the product purchases and the time and type of marketing contacts. That data would typically come from marketing and sales data.

4

USE CASE

NEXT BEST OFFER

Next best offer allows an organization to increase its up-sell and cross-sell opportunities by predicting what the customer wants next.

This is achieved by analysing the customer's market basket and finding the patterns between products to forecast future purchases. Leveraging this knowledge can ultimately lead to improved marketing ROI, customer loyalty and sales results.

TOP USES
FOR NEXT BEST OFFER

4.1

ENHANCING LOYALTY THROUGH NEXT BEST OFFER.

Offering a customer an additional product that suits his or her wants and needs will not only increase sales it can also strengthen customer relationship. Relevant offers might increase the interest and "customer stickiness" to banking products.

PAGE 35

4.2

MEASURING PRODUCT PROPENSITY

Product propensity scoring can be used to boost revenue by offering customers products or services that they really want.

PAGE 37

4.3

PRODUCT BUNDLING TO UPLIFT REVENUE

Determining which products are most likely to be purchased together is essential in order to develop highly effective product promotions and to increase revenue.

PAGE 39

ENHANCING LOYALTY THROUGH NEXT BEST OFFER

Next best offer can find the products customers want and when they want it, which can enhance customer loyalty and increase sales.

Next best offer is commonly used to re-engage with customers, reminding them of the brand or service while simultaneously opening cross- and up-selling opportunities. Picking the wrong products to offer the customer can cause a disjointed feeling between customer and organization or perhaps worse, the marketing messages being perceived as spam. Generating personalized offers in a timely fashion can help drive loyalty and increase ROI from existing customers.

TESCO

Tesco has focused its Next best offer strategy on increasing sales to regular customers and enhancing loyalty with targeted coupon offers. Shoppers who buy diapers for the first time at a Tesco store are mailed coupons not only for baby wipes and toys but also for beer, since Data analysis revealed that new dads tend to buy more beer, because they are spending less time at the pub.

NETFLIX

Netflix is a completely data driven company and can accurately predict what movies its customers like and hate. To achieve this, Netflix feeds all its customer data into a recommendation system that takes into consideration all aspects about the customer, from viewing behaviours to customer profiling. Netflix then use predictions generated from the recommendation engine to make tailored movie recommendations based on each customer's preferences.

HOW

Linking the customers market basket or historical transactions to products they may want to buy in conjunction with past purchases is often achieved through Market basket analysis. Market basket analysis employs the link analysis technique where links obvious or otherwise are found between products. Once links are found timing is also key as some products are highly time sensitive. For example, customers may only take out student bank accounts and related products once in their lifecycle and targeting marketing efforts towards these customers to soon or too late can cause banks to miss an opportunity. Decision trees are particularly useful in this regards as they can be used to score suitable timings for a marketing messaging, based on past customer experiences.

ALGORITHMS*

*See page 52 for Appendix

DATA

Transaction history of the customer is the most important data for this use case. Profile data is also required to help predict the customers life stage events to optimize the timing of marketing messages.

MEASURING
PRODUCT PROSPENSITY

Identifying the products or services customers are most likely to be interested in for their next purchase is a valuable piece of information for any marketing managers.

Measuring a customers product propensity involves scoring each product or service based on the customers peers, historical purchases and behavioural characteristics. This scoring can then be used to produce a product or service ranking, with the highest ranking product or service being the next best offer for the customer.

Westpac, a bank operating in Australia and New Zealand with 812,000 customers, are successfully making use of next best offer to drive their cross sales. Using the next best offer methodology, Westpac measures each customers product propensity across their range of products and services. From this information, Westpac is able to subscribe extra banking products to 37% of its customers through its branch staff and 60% to this customers who communicate through its call centre staff.

A pioneer in Next best offer is Amazon. The company employs collaborative filtering to predict a customer's product propensity based on the customer's peers. From this, Amazon generates its "you might also want" and "customers also bought" campaigns for each product purchased and viewed by the customer. This has led to a significant increase in Amazon's cross sales and revenue.

HOW

A customer's propensity towards a certain product is an interesting case as it involves using historical data to score the likelihood of future purchasing habits. Scoring for product propensity is typically achieved through a classification algorithm such as linear regression and decision trees. Decision trees can help to find significant attributes that companies want to consider in their propensity calculation. This may include the customers income or demographics. The top attributes can then be uncorrelated using a method such as PCA and used as the inputs for the linear regression algorithm. Linear regression can then simply score customers propensity for each product as a percentage.

ALGORITHMS*

*See page 52 for Appendix

DATA

The customer's propensity towards a certain product can be measured through both their own historical purchases and their peers. This requires customer profile data and their transaction history.

PRODUCT BUNDLING
TO UPLIFT REVENUE

Bundling products that go well together can differentiate an organization from its competitors. The banking world can apply product bundling to increase cross– and upselling in a better way.

For banks to move forward, they need to offer tailored products or bundles of products and services.

On average, people have three to four banks with two or more products in each bank. Using bundles, based on preferences, banks can shift customer utilization and revenue significantly.

Manulife Financial

Manulife Financial makes use of data analytics to bundle the deposit and credit account. As Canada's first flexible mortgage account, the bank offers a mortgage with checking and savings, using the net balance to calculate the specific interest – which positions the account to serve as the client's primary current account.

eircom

Eircom is an Irish provider of telecommunications services that makes use of predictive models based on customer behaviour information. The predictive modelling enables Eircom to find the most popular bundles to offer their customers and allows product development to use this knowledge to find more innovative ways to obtain new customers and retain existing ones.

HOW

Finding which products naturally go well together has been a challenge for anyone wishing to develop an effective cross- and up-selling campaign. Historically link analysis has proved extremely effective at finding products that sell together, even if the link at first may be obscure or non-obvious. This technique is widely used in retail to aid in product placement. Nearest neighbour is also a popular approach to this problem as it follows the logic: similar customers will purchase similar products. A pitfall of this approach is the dynamically changing situation of the customer, who may look very different from the time they bought the product. Decision trees are a strong candidate for cross-selling problems as they can show the attributes of the customers who buy certain products. This perhaps makes them the most useful in this particular use case.

DATA

Customer profile and transaction data is needed to build an effective cross selling model. Care should be taken as many products in financial services are bundled with other products or have been bundled in the past. This creates a situation where some product data can be synonymous with entirely different product data, so extreme care and business knowledge is required to find these cases.

5

USE CASE

CHANNEL JOURNEY

With so many ways a customer can interact with a company, including mobile, social media, clickable ads, stores, TV and publication platforms, keeping track of the customer journey can be a difficult task.

Big Data helps by taking a holistic view of the entire customer journey and experiences on each channel. This can be used to find patterns of usage that lead to sales, or where and which channels are underperforming. From this knowledge, banks can optimize the funnel conversion, increase messaging effectiveness and measure marketing results across all channels.

TOP USES
FOR CHANNEL JOURNEY

5.1

PROVIDE MORE RELEVANT CONTENT IN THE PREFERRED CHANNEL

Customers may not use all channels in the same way and so messaging on each channel should reflect this. Some segments may prefer to research products on a blog before visiting the website and landing pages could adapt to this by displaying relevant offers and further reading linked to their research.

PAGE 43

5.2

RECOGNISE MULTI-CHANNEL BEHAVIOURS THAT LEAD TO SALES

Big data analytics allows companies to recognising patterns of channel usage in their customer segments and how successful journeys lead to sales and where unsuccessful channel journeys fail. This then allows to optimize the funnel conversion by optimizing bottlenecks and fixing points of high drop off rates in the customer channel journey.

PAGE 45

5.3

GUIDE CUSTOMERS TO LOW COST CHANNELS

Sometimes different channels can provide the same service, but at widely different costs to a company. Aided by Big Data analytics, guiding customers to low cost channels requires robust self- service and a customer driven approach to provide a quality and intuitive service.

PAGE 47

5.4

MEASURE MARKETING EFFECTIVENESS ACROSS CHANNELS

The customer may purchase a product in one channel, but had made the decision to purchase it because they had seen an advertisement in another channel the week before. The question is how much credit should be given to each channel for selling the product?

PAGE 49

PROVIDE MORE RELEVANT CONTENT
IN THE PREFERRED CHANNEL

Sending the right message through the right channel can be just as important as the message itself.

Big data analytics can be used to accurately find which channels customers are using and how they are using them. This knowledge can form a strategic advantage to reach specific target segments in the most cost effective way and maximising marketing budgets. How customers are using channels may also highlight areas where banks need to focus resources, such as which channels are used for complaints or which are primarily used for research.

HDFC BANK
We understand your world

HDFC bank is the 5th largest bank in India by assets with an estimated value of £39 billion. HDFC uses their knowledge of the channel journey to personalize the customers experience on their channel of choice. For example, an ATM would recognise the customer's preferred language from previous interaction on the organization's website. Coupled with other preferences measured across a wide range of channels has decreased time at ATMs by 40% and helped reduce operational costs.

OCBC Bank

OCBC bank is headquartered in Singapore and has assets estimated to be over £59 billion. OCBC identified a significant market segment moving to online channels and the expectations of these channels are growing with rapid pace of technology. One of OCBC's core principles is to deliver a high quality service through the channel of choice, by adapting to what the costumer expects.

中国银行
BANK OF CHINA

Bank of China has created a new online banking platform to examine data from various channels accessed by its global customer base. The online platform integrates customer-facing systems such as branch, phone, mobile and web services for Bank of China's 100 million customers. This enables Bank of China to provide the right content at the right channel.

HOW

Measuring the customer's channel of choice can be achieved by building a profile of each customer based around the frequency of interaction on each of the channels. The type of contact can also be measured either through engagement metrics or text analytics. For example, text analytics can be used to measure if a person has made a complaint, while analytics from URL referrals and cookie data can assess if somenone is researching the product or is a frequently returning customer. Decision trees can also be employed to find the attributes of customers who use particular channels and what purpose they use them for. This information may aid in future segmentation and in the design of the channels.

ALGORITHMS*

*See page 52 for Appendix

DATA

Data from online usage from cookies, URL referrals and metrics can be used to determine which online channels the customers are using and what they are using them for.

RECOGNIZE MULTI-CHANNEL BEHAVIOURS

THAT LEAD TO SALES

Knowing how customers use different channels can help financial institutions to optimize them and to increase sales.

Customers may use multiple channels to both research and purchase banking products. Understanding these patterns in channel behaviours can help uncover which medium is generating sales either indirectly through customer research or directly through sales.

This knowledge can be used to help optimise funnel conversion for both lead and sales management and also be used to locate which channels are over and underperforming.

vodafone™

Vodafone recognised their traditional approach of measuring the customer channel journey using cookies was adversely affecting their "return to basket" email campaign. This was due to 40% of their customer base using their phones for emails, which use different cookies than their laptops, so information about their abandoned shopping basket was lost between channels. To solve this, Vodafone deployed deep linking data analytics to track the customer across channels and platforms and they now deliver a completely cookie-free experience. Return to abandoned baskets has since increased by 30%.

GE Capital

GE Capital is a financial service branch of the American conglomerate General Electric. GE Capital understands that people are omni-channel users and that their journey to GE Capital's products can vary greatly from person to person, but each one expects a smooth transition from channel to channel. By understanding how customers migrate from different channels, GE Capital has optimized its customer's experience across online and offline channels, increasing both customer satisfaction and marketing efficiency.

MoneySupermarket.com™

MoneySupermarket measures offline and online user interactions across a range of channels looking for its interest in holiday, car hire and hotels to target customers for travel insurance. By combining data across all channels to build a customer profile has led to a 33% increase in attribute revenue.

HOW

Tracking the user interaction across the diversity of channels is a challenging task. Metrics such as customer engagement on each channel linked to the customer profile should give an indication on which of the segments are using each channel.
Other metrics such as navigation page to content page ratio should also indicate the efficiency of the online and mobile channels.

DATA

Customer profile data will be required to view which and how customer segments use channels. Offline metrics such as channel performance, time stamps and any comments made about the customers visit can be used to build a picture of their offline journey, while online metrics gained from cookies, webpage JavaScript, tracking codes and Matchback are used to build the online journey.

Tracking customer between online channels is typically achieved through cookies, URL referrals and matchback. Measuring the influence of offline media can be achieved directly by using promotional codes and campaign specific tags.

These however, are not always available, so A/B testing and measuring lift associated with an offline campaign are often used by multichannel marketing managers. The type of interaction types across the channels can be measured using a combination of text analytics and decision tree classification.

ALGORITHMS*

*See page 52 for Appendix

GUIDE CUSTOMERS TO LOW COST CHANNELS

Assisted self service on online channels is facilitating to migrate customers away from traditional brick and mortar stores.

Replacing high cost channels with low cost alternatives is helping reduce costs and increase organisational efficiency. Aiding customers during this channel migration has put increased pressure on assisted self-service and delivering a constant experience across the multi-channel market. Part of the drive towards a consistent and intuitive service is being helped by Big Data analytics measuring the effectiveness of the channels and applying techniques to increase self- service efficiency.

HSBC

HSBC found the primary barriers for internet banking were customer habit, security concerns and a lack of confidence. They now have an active migration strategy to address these concerns. Part of the HSBC migration strategy is to enable customers to undertake increasingly more complicated banking activities via the internet.

UniCredit

UniCredit transaction system was originally developed with an internal employee focus and later re- developed with a consumer online centric approach. This enabled UniCredit to retain customers and increase their volumes by 25-50%, by guiding transactions to lower cost channels.

HOW

Converting customers to low cost channels must ensure those are intuitive, but can also handle a range of complex banking activities. To understand how good the ways of distribution are at this, requires metrics and classification algorithms to score the effectiveness of the channel for each segment. Metrics such as engagement, content/navigation scores, drop off rates, time-stamps and number of clicks to action can all help measure how effective online mediums are. Classification algorithms such as a decision tree using these metrics can then help score the entire online experience and find areas of the channels that need addressing. For example, links to information pages for a product may be situated off screen and a score based system may show people who visited these pages are not as well informed as other products.

ALGORITHMS*

*See page 52 for Appendix

DATA

Data from online JavaScript and time stamps can indicate the kind of content and how long the user has been viewing the page. This is then combined with the profile data to determine how each customer is using the channel.

MEASURE MARKETING EFFECTIVENESS
ACROSS CHANNELS

Being able to track and measure marketing effectiveness across channels allows marketing managers to optimize their budgets delivering higher ROI on their marketing campaigns.

Identifying the value of marketing across different media usage is a serious challenge for a business to improve their marketing campaigns in a rapidly changing environment. So data analytics can be used to improve the gut feeling approach when it comes to marketing measurement.

Bank Polski

PKO Bank Polski, a leading banking institution in Poland, employs a multi-channel campaign management platform. By using this platform, the bank can perform analytics across its channels to develop a clearer picture of its overall marketing effectiveness leading to a higher ROI.

LAURENTIAN BANK

Laurentian Bank of Canada uses data analytics to aid its understanding of the performance, factors and influence of its marketing campaigns across its various channels. This allows Laurentian Bank marketing managers to appropriately change its campaigns accordingly when they are not performing well.

HOW

Measuring campaign effectiveness across the different mediums requires various metrics to determine offline-online and online-offline interactions, this is typically achieved by measuring associated lift and ROI lift. Decision trees can then be used to score each channel, indicating its performance and how many indirect/direct sales the channel generates against the cost of the channel. Text analytics can also be used to measure the types of social sphere interactions, which are classified by the decision tree.
subsequently

ALGORITHMS*

*See page 52 for Appendix

DATA

Measuring online click streams, tracking codes and cookie data is required to build a complete picture of the user's online journey. This is then combined with the customer profile and Ad-statistics from all the mediums used to build a picture of channel performance.

APPENDIX

ALGORITHMS

Data is the lifeblood of the Big Data use cases discussed throughout this white paper, but equally algorithms are the organs turning data into insights and value. Both are equally needed to turn these Big Data use cases from theory into reality. This next chapter discusses the algorithms that are used to turn data into business insights.

ALGORITHM
INDEX

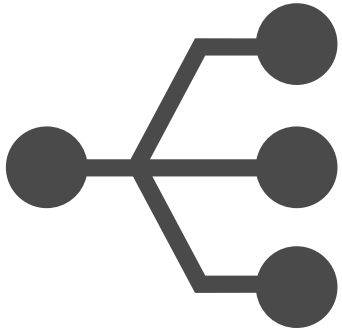
ALGORITHMS
DECISION TREES & RANDOM FORESTS

Decision trees are one the most powerful data mining techniques, as they can handle a diverse array of problems. One of their core strengths comes from their ability to handle almost any data type, numeric, nominal or otherwise.

Decision trees work by splitting data up along its dimensions into smaller data cells, with the aim of decreasing the overall entropy of the data within each cell. Each cell is then treated independently and is further split until a target entropy level is reached. To maximise the efficiency of this process, the decision tree tests all variables to find the most important variable that decreases the entropy by the largest amount in the cells after the split.

A typical decision tree would be used to classify data into a predefined target field. An example field might be high probability of churn or even a numeric value such as an estimated price in the case of regression trees. Decision trees are also used to explore which are the most important fields for a particular data set, which then can be used in other algorithms.

Random forests is a technique used to boost the efficiency of decision trees models by creating an ensemble of slightly varying decision trees that model the same target. By having slight variances between the decision trees acts as a safety net between possible errors and noise of an individual decision tree. Each tree in the ensemble then votes a target field, where the winning target field is then assigned to the data.

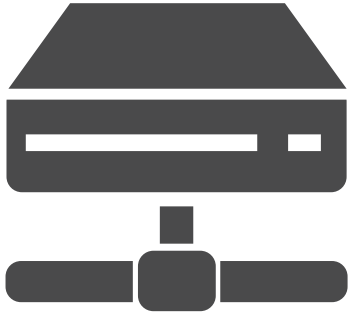


ALGORITHMS
CLUSTERING

Cluster detection is simply the automation of finding meaningful patterns within a data set. Often the problem in data mining is not a lack of patterns but identifying which patterns are useful amongst the noise of possibly hundreds of competing patterns within the data. Cluster detection helps penetrate this noise by finding clusters of data that form natural groupings within the data set. For example, one such cluster might contain active people who have a low risk of heart disease, or trees with needles do not lose their leafs in winter. Clusters essentially breaks a complex problem up into much smaller manageable pieces where a human can then easily identify the underlying pattern. After clusters have been detected, they are often themselves the object of study.

Cluster detection is often employed as an undirected technique as there is no target variable. This is unlike the other algorithms discussed in the chapter, which prescribe a target variable to the data. Although they are undirected, cluster detection is often used to solve directed problems, because clusters are sought to solve some business purpose. In marketing clusters are nearly always called segments, which is a popular application of the clustering technique.

K – means is probably the most popular clustering algorithm used to find segments within a data set. K-means works by assigning a K number cluster seeds randomly with the data's dimensions. Data points are assigned to the nearest cluster seed, which is then moved to the average position of the data points. This process is carried out iteratively until convergence.



ALGORITHMS
TEXT ANALYTICS

With the amount of text being generated in organizations, both internally on reports or CV's and online in the social media sphere and news reports it has become humanly impossible to read all these documents. Text analytic algorithms assist by automating the reading process and providing a brief summary compiled from possibly thousands of documents. Text analytics algorithms heavily rely on probability theory and the rarity and occurrence of certain words, which can be used to predict the meanings and themes of the text.

Text analytic techniques are a form of classification algorithm so a target field for the algorithm is clearly defined. In sentiment analysis, typical target fields would be negative, neutral and positive sentiment. As text analytics are probabilistic models the output field can be scored, for sentiment analysis the scoring is typically between -1 (negative sentiment) to +1 (positive sentiment). Another example of target fields would be to classify news articles on whether you expect them to have or not to have an influence on the stock market. Naïve Bayes is perhaps the popular algorithm for text analytics. Bayes formula was made famous for its use in determining the probability of medical tests false positive and false negatives rates.

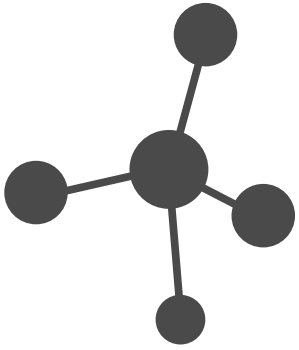
Naïve Bayes is a slight modification of Bayes formula to determine the probability that certain words belong to documents of particular classes. For example, the word Viagra is heavily associated with spam emails and today naïve Bayes is widely used in identifying spam emails and is used in almost every email checker to date. The naïve part of naïve Bayes come from the fact the algorithm treats each word independently and document is simply viewed as a bag of words. Wider context of the words is then subsequently lost using naïve Bayes.

ALGORITHMS
NEURAL NETWORKS

Artificial neural networks form a powerful class of classification algorithm that can be used for a wide range of problems. They are based on the neural networks found in biology and the human brain, where nodes are activated by a signal that in turn transmits a response signal to activate other nodes. In artificial neural networks, each node is comprised of a combination function that receives various incoming signals and calculates the total received signal based on a set of weights.

A transfer function then outputs a signal based on the total revived signal from the combination function. Neural networks are designed to work with numerical values that are approximately within the same range. For example, using the fields wage $65,000 and age 45Yr, the wage field would heavily basis the algorithm as wage orders of magnitude higher than age. This can be corrected by adjusting the units so wage could become $65K, which is only 20 "units" away from 45Yr.

Neural networks are a classification type algorithm, which means they assign data to a predefined target field. The target field would typically be a scoring function, such as the customer's propensity towards a particular product, or an estimated value for a house. The algorithms are, however, considered a black box, as they cannot tell you why a certain result was produced. This can restrict the use of neural networks when the question is why does person/ event A lead to action B, but they are exceptionally powerful in answering what is the most likely action B given person/event A.

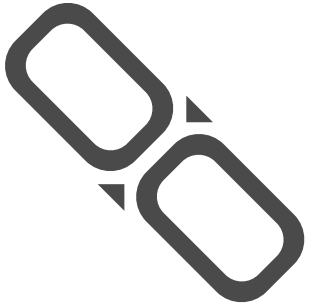


ALGORITHMS
LINK ANALYSIS

Link analysis was built around the need to find relationships and connections in this ever-more-connected world. It answers questions such as who is friends with who on Facebook, which drugs are being prescribed to which patients, who is calling whom on the telephone and which webpages bridge language barriers. The answers to these and many other related questions are all contained within the data, where Link analysis is able to take direct advantage.

Link analysis is part of a subset of mathematics called graph theory, which represents the relationship between objects as edges and the objects themselves as nodes. Link analysis falls under the category of both direct and undirected data mining. A popular use of link analysis is to create new variables to be used by other modelling techniques, but link analysis can also be used in an undirected way by exploring the properties of the graph itself.

Areas where link analysis has been particularly successful include identifying key sources of information on the web by analysing links between pages, finding influential customers from call pattern data to recruit new subscribers from competing networks and understand physician referral patterns to gain insight into insurance fraud. In some cases the links between objects are explicitly recorded, such as HTML encoding between webpages or call logs on telephones. These sources of information, however, can take significant data preparation before they become useful. Other sources of information may only imply an implicit link between objects, in these cases the data scientists challenge is to identify the links.



ALGORITHMS

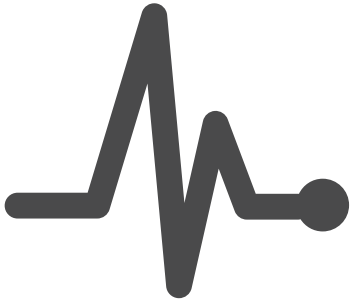
SURVIVAL ANALYSIS

Survival analysis, sometimes called time-to-event analysis is a technique to explain when you should start worrying about an event. Survival analysis was born from studying survival rates of hospital patients and manufacturing failures, but used for marketing it can tell you when you can expect a customer to leave your service or stop buying your products. By providing answers to these questions, survival analysis provides snapshots of the customer's product lifecycle and tell you when you should start to worry about your customer relationship. Survival analysis has historically been a powerful tool for answering the following questions:

- When is a customer likely to leave?
- When will the customer move to a new customer segment?
- The next time a customer will narrow or widen their interaction with the organization
- Which factors are likely to increase or decrease customer tenure?
- Quantitative effects of various factors on the customer tenure

The insights gained from this understanding can be directly used in the marketing process and help determine how long groups of customer will stay with the organization and therefore how profitable a particular customer segment will be.

Survival analysis is calculated using survival curves and hazard probabilities, where databases of tens of thousands to millions of customers are used in the calculation. Every survival curve starts at 1 and decreases to 0 as customers begin to leave the organization. Often the curve will never reach 0 as some people with long tenures will always remain active. Survival curves of different customer groups then provide a direct way of comparing the under and over performance of different segments, indicating that segments value to the organization.



METHODOLOGY

Our research began by finding hundreds of industry examples across the financial and related sectors to build an overview of what Big Data could do for banks. We as the Big Data Research Team at EVRY then grouped the industry examples and reviewed them in our team. Patterns soon started to emerge, forming groups and strategies of Big Data methods that can be applied by marketers in the banking sector. Based on our profound research we could extract the information and value out of the various sources to build the many comprehensive use cases that can be found in this whitepaper.

CONDUCTING

SECONDARY RESEARCH

Finding industry examples of Big Data across the financial and related sectors

- Conference notes
- Industry studies
- Academic books & articles
- Vendors commercial information
- Internet articles

FILTERING

INDUSTRY EXAMPLES

Grouping of use cases

INTERNAL REVIEW

IN BIG DATA TEAM

Forming comprehensive use cases

- Extracting business value
- Prioritize most important usage
- Outlining algorithms
- Informing on data sources

WHY EVRY

When it comes to Big Data within the Norwegian banking market EVRY is placed in a unique position. EVRY has extensive expertise in handling Norwegian banking data and understands the unique demands of the Norwegian banking culture. We not only have the internal resources and competencies to develop the right solutions, EVRY is the only provider who has the access to and expertise of Nordic Bank Inhouse and public data to deliver truly Big Data solutions. We also take researching Big Data in banking seriously, identifying the problems facing banks today and finding workable Big Data solutions. This research forms the cornerstone of EVRY's Big Data projects, but we are not just researching Big Data, EVRY is currently developing and experimenting on several practical Big Data solutions. Combining this research and practical knowledge enables EVRY to not only advise on which Big Data solutions are best for the banks, but also recommend how the banks might deploy their solutions and what pitfalls and challenges they might face.

For more information, please contact us.

AUTHORS
OF THIS WHITE PAPER

YOOK-PEI SHEE

INNOVATION MANAGER
FINANCIAL SERVICES

YOOK-PEI.SEE@EVRY.COM
+47 9220 5704

DAVID CROMPTON

DATA SCIENTIST
BIG DATA RESEARCH

DAVID.CROMPTON@EVRY.COM
+47 9061 9155

HILLE RICHTER

BUSINESS RESEARCH
BIG DATA RESEARCH

HILLE.RICHTER@EVRY.COM
+47 9809 1342

SVENN-PETTER MÆHLE

INTERACTION DESIGN
BIG DATA RESEARCH

SVENN-PETTER.MAEHLE@EVRY.COM
+47 9443 4706

SOURCES

AMAZON
Kyathitechnologies (n/d): http://www.portlandps.vic.edu.au/pdf_upload/server/php/58/ files/28186e4f385ea35786ca4ef4b7bfcca0e2f3.pdf

AMERICAN EXPRESS
Datamashup (2014): http://www.datamashup.info/ten-big-data-case-studies-in-a-nutshell/

BANK AUSTRIA
(n/d) https://www.teradata.com/case-studies/BAWAG-PSK-CRM-EB6272/?type=CS

BANK OF AMERICA
Gordon S. Linoff, Michael J. A. Berry (2011): Data Mining Techniques, PP 11

BANK OF AMERICA
Forbes: Tom Groenfeldt (2013): http://www.forbes.com/sites/tomgroenfeldt/2013/06/11/banks-use-big-data-to-understand- customers-across-channels/

BANK OF CHINA
IBM (2014): http://www-03.ibm.com/press/us/en/pressrelease/42845.wss

BARCLAYS
NEPIC (2009): http://www.nepic.co.uk/BASME/docs/Barclays_case_study_Feb_2014.pdf

BARCLAYS
Our Social Times (2014): http://oursocialtimes.com/how-to-use-social-media-monitoring-for-a-product-launch/

BBV
American Banker: Sean Sposito (2012): http://www.americanbanker.com/btn/25_8/bbva-citi-use-artificial-intelligence-to-understand-customers-1051176-1.html

BBVA
IBM (n/d): http://www-01.ibm.com/common/ssi/cgi-bin/ssialias? infotype=SA&subtype=WH&appname=SWGE_YT_YT_USEN&htmlfid=YTC03367ESEN&attachment=YTC03367ESEN.PDF

CAESARS ENTERTAINMENT
Capgemini Consulting (2012): http://www.capgemini-consulting.com/resource-file-access/resource/pdf/caesarsentertainment_9thapril.pdf

CITIBANK
Economist (2012): http://www.economist.com/node/21554743

DELL
GNIP: Amanda Bridenhagen (2013): http://blog.gnip.com/tag/sentiment-analysis/

ETHIAD
Big Data Startups (n/d): http://www.bigdata-startups.com/BigData-startup/etihad-airways-big-data-reach-destination/

EIRCOM
Portraitsoftware (n/d): http://www.portraitsoftware.com/sites/portrait/files/media/pdf/case -studies/eircom-case-study.pdf

FIFTH THIRD BANK
Acctiva (2014): http://acctiva.co.uk/news-media/news/Banks-Set-Stage-For-Customer-Acquisition-With-Data-Analytics/

FIRST TENNESSEE BANK
IBM (2011): http://www.ibm.com/smarterplanet/global/files/sweden__none__banking__First_Tennessee.pdf

FIRST TENNESSEE BANK
Nielsen (2009): http://www.nielsen.com/content/dam/corporate/us/en/newswire/uploads/2009/06/segmentation -and-customer-loyalty-white-paper.pdf

FORD FIESTA
McKinsey (2012): http://www.mckinsey.com/insights/marketing_sales/demystifying_social_media

GE CAPITAL
Business Wire (2013): http://www.businesswire.com/news/home/20131125005618/en/GE-Capital-Retail-Bank-Study-Reveals-
Rise#.U4wuXPmSxB0

HDFC
NCR (2010): http://www.ncr.com/newsroom/resources/hdfc-bank-and-ncr-jointly-win-two-asian-banker-technology-implementation-awards

HDFC
Mthink (2006): http://mthink.com/article/hdfc-bank-enhances-bottom-line-revenue-through-customer-lifecycle-marketing/ [HMV] SAP (n/d): http://www.kxen.com/Customers/HMV

HSBC
EFMA (2009): Microsoft, Innovation in multi-channel management, PP 26

INTUIT
Elisabeth Horwitt (2011): http://searchcio.techtarget.com/news/2240036281/Sentiment-analysis-turns-social-media-into-customer-intelligence

JEANSWEST
Marketing (2013): http://www.marketingmag.com.au/case-studies/case-study-jeanswest-from-batch-and-blast-to-targets-and-triggers-
42604/#.U4RvUaM4WUk

KIA
Bloomberg Business Week: Rachel King (2011): http://www.businessweek.com/stories/2011-03-01/sentiment-analysis-gives-companies-insight- into-consumer-opinionbusinessweek-business-news-stock-market-and-financial-advice

LAURENTIAN BANK OF CANADA
IBM (n/d): http://www.bankingreview.nl/download/25887

MANULIFE FINANCIAL
McKinsey & Company (2012): http://www.mckinsey.com/App_Media/Reports/Financial_Services/
Retail_Banking2010_Relationship.pdf

MONEYSUPERMARKET
BlueKai (2013): http://bluekai.com/case-studies.php

NATIONWIDE
Pitney Bowes (2011): https://www.pb.com/docs/US/Products-Services/Software/Customer-Strategy/Real-Time-Customer-Decisioning/ Portrait-Interaction-Optimizer/PDFs/Nationwide-case-study.pdf

NEDBANK
Darryl K. Taft (2014): http://www.eweek.com/database/ibm-rides-analytics-big-data-to-banking-wins.html

NETFLIX
Ravi Kalakota (2012): http://practicalanalytics.wordpress.com/2012/01/05/analytics-case-study-schwan-foods/

OCBC
Adam Ramshaw (n/d): http://www.slideshare.net/aramshaw/implementing-triggerbasedmarketingtodrivecustomerloyalty

PKO BANK POLSKI
Teradata (2012): http://www.teradata.com/News-Releases/2012/PKO-Bank-Polski-and-Teradata-Deploy-Multi-Channel-Campaign-Management-Platform/ Digital Alchemy

ROYAL BANK OF CANADA

SOURCES

Banktech (2013): http://www.banktech.com/business-intelligence/how-financial-institutions-can-use-big-d/240155874

RYANAIR
AT Internet (2011): http://www.atinternet.com/wp-content/uploads/2012/02/AT_CS_Ryanair_EN.pdf, Sascha Mayer (2007) Ryanair
and its low cost flights in Europe

STARBUCKS
McKinsey (2012): http://www.mckinsey.com/insights/marketing_sales/demystifying_social_media

TATRABANK
SAS (2012): http://www.sas.com/offices/europe/norway/nyheter/tatrabanka.html

TESCO
Thomas H. Davenport, Leandro Dalle Mule, John Lucker: Harvard Business Review (2011): http://hbr.org/2011/12/know-what-your- customers-want-before-they-do/ar/1

T-MOBILE USA
Big Data Startups (n/d): http://www.bigdata-startups.com/BigData-startup/t-mobile-usa-cuts-downs-churn-rate-with-big-data/

UNICREDIT
EFMA (2009): Microsoft, Innovation in multi-channel management, PP 27

VODAFONE
Experian (2013): http://www.experian.co.uk/assets/marketing-services/case-studies/case-study-vodafone.pdf

WESTPAC
Stuart Corner (2014): http://www.smh.com.au/it-pro/business-it/westpac-using-big-data-to-woo-customers-with-offers-made-to- measure-20140303-hvfx5.html

ZIONSBANK
Bank Technology News (2013): http://www.americanbanker.com/issues/178_190/zions-bank-combs-big-data-for-customer-preference-clues-1062531-1.html

Akin Arikan (2008): Multi Channel Marketing: Metrics and Methods for On and Offline Success
Banking Strategies: Sherief Meleis & Rich Solomon (2012): http://www.bai.org/bankingstrategies/product-management/pricing/segment-based-
deposit-pricing
Big Data Start-ups (n/d): http://www.bigdata-startups.com/best-practices/
Business Case Studies (2014): http://businesscasestudies.co.uk/barclays/discovering-customer-needs-through-research/ introduction.html#axzz32zZTHLAl
Cathy O'Neil; Rachel Schutt (2013): Doing Data Science
Computer weekly: Bill Goodwin (2013): http://www.computerweekly.com/news/2240179949/Google-mines-data-to-identify-the-best-managers
EVOLV (2014): http://www.evolv.net/success-stories/
IBM (2014): http://www.ibm.com/Search/?q=Case+study&v=16&en=utf&lang=en&cc=uk&Search=Search
Business Insider: Jeff Dachis (2012): http://www.businessinsider.com/big-data-is-the-future-of-marketing-2012-7
Our Social Times (2014): http://oursocialtimes.com/how-to-use-social-media-monitoring-for-a-product-launch/
SAP (n/d): http://www.saphana.com/community/implement/use-cases/content?filterID=contentstatus%5Bpublished%5D~category% 5Bbanking%5D
The Financial Brand (2014): http://thefinancialbrand.com/32653/big-banks-social-media-sentiment/

We bring information to life

WWW.EVRY.COM