DYNAMIC CLUSTER BASED MARKOV MODEL FOR DEMAND FORECASTING

Albert F. H. M. Lechner, Steve R. Gunn

School of Electronics and Computer Science University of Southampton, UK

ABSTRACT

Sales forecasts are essential to every business strategic plans and can both save the business money and increase its competitive advantage. However, many current businesses underestimate the opportunities accurate forecasts provide and rely on judgemental forecasts from experts within the business. Machine learning and statistical forecasting methods are used by both academics and practitioners to increase the accuracy of these forecasting methods and can be further improved by applying the newly developed dynamic cluster based Markov model, presented in this work. This approach gathers global sales pipeline data to build a short-term sales forecast. The prediction of future sales for the next three months is improved over a regular Markov transition model. The new model can support short-term planning, thereby supporting regional and product-specific forecasting to steer business activities to their given targets and remain profitable.

KEYWORDS

Demand forecast, Time series data, Clustering, Markov model

1. INTRODUCTION

Artificial Intelligence, or AI, has become increasingly important in everyday life and is implemented in more daily applications, such as online searches, voice recognition, face detection, and traffic sign detection [1]. Recently, machine learning (ML) algorithms have managed to beat the world champions in the game GO, which is considered the most difficult game for a computer to beat as the possible moves are nearly unlimited [2]. ML is one of the most prominent tools of AI and carries out tasks in a way humans consider smart. In ML machines are given specific data and learn for themselves [3]. Recently, self-driving cars have become not only a highly discussed topic, in which ML algorithms decide on the car's actions without a physical driver acting, but they are also a highly researched area [4, 5, 6]. However, self-driving cars are not the only application of ML for manufacturers. Current research in that area does not only focus on autonomous driving, it also expands to all areas of every businesses value chain. As more data and computational power becomes available, ML is used in many parts of a manufacturers value chain, such as development, procurement, logistics, production, marketing, sales, and after-sales, followed by a connected customer after the purchase [7]. Analysing past and current data to improve business is an important task, but predicting the future is even more important as several of a company's decision-making processes are based on forecasts. Important decisions such as strategic planning, production planning, sales budgeting, marketing planning or

new product launches are influenced by forecasts. Therefore, many practitioners and researchers have focused on new forecasting methods and improved forecasting accuracy as money can be saved and a business's competitive advantage could be improved [8, 9]. Emerging ML models have shown increasing accuracy but they have challenges and shortcomings from a practical aspect as they typically have a dense black box structure which often make them difficult to explain them within a business environment. It is important for businesses to not only increase the accuracy of a forecast, but also to focus on explainability for the wider business network. Understanding the main features that drive the prediction is not only important for forecasting but also essential for other departments to affect the important features and thereby increase future sales [10]. Forecasting the buying process of a potential future customer supports short term planning on a regional as well as product specific basis. Various techniques are able to model this process but suffer from drawbacks being applied on real-world problems. To overcome these problems a new approach was developed based on a Markov model. So far Markov models have not been combined with a classification that has moving boundaries based on seasonal patterns for increased sales forecasting accuracy as well as easier business interpretability. Such problems are faced by many companies in the automotive industry and a case study is provided in this paper.

2. RELATED WORK

This section provides an overview of techniques used for forecasting. Although different academics and practitioners use these techniques to improve forecasting, this work identifies a new approach that improves forecasting accuracy based on the methods presented. The forecasting methods introduced range from classification algorithms to time series methods, like ARIMA models. The classification, using neural networks, is used in this research to separate different customer groups to separate a single time series into different ones which share similar buying behaviour. Since they are appropriate for forecasting time series, ARIMA models are introduced and used to separate clusters. These new findings support sales and demand forecasting for a variety of different businesses.

2.1. Stage Transitions

Modelling transitions between stages and individuals transitions within a given framework, like patients moving from the stage Alive to the stage Death is a widely applied research topic. Patient transitions in a medical environment is the most common research topic within the multi-state modelling research, but other fields are researched within similar frameworks [11, 12, 13, 14]. For example [15] used a Markov model to forecast educational enrolments over time. [16] used multivariate Markov chain models for demand prediction, focusing on customers who already had a large individual sales history within a company's database. The transition between the two states can be measured with a hazard rate [11]. By adding more stages to the two-stage model, such as Unwell, the standard survival model needs to be extended with a multi-state model (MSM). An MSM is a time stochastic process, which allows individuals to move between finite numbers of states [17]. The different stages can be transient or absorbing, depending on whether a transition could emerge from the state. The transition between the stages will often be incomplete due to left-or right-censored observation times, where the event occurred before the study started or for right-censored when the study ends before an event occurs. Incomplete data occurs when the process is not observed from the origin or, for the right-censored case, if the individual is not in an absorbing stage when the data was captured.

2.2. Markov Models

Markov models are frequently used because of their simplicity. A Markov chain is a sequence of random variables (X_1, X_2, X_3) where the probability from moving from one state to the other only depends on the previous state, the so called Markov property. The Markov property is formulated as [18]

$$p(X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \dots, X_0 = s_0) = p(X_{t+1} = s \mid X_t = s_t)$$

for all times $t = 1, 2, 3, \dots$ and for all states $s_0, s_1, \dots, s_t, s$. In Markov models, the transition probabilities can be calculated from the intensities by solving the forward Kolmogorov differential equation [19]. Markov models can be differentiated into time-homogeneous models and non-time homogeneous models. In time homogeneous Markov models, all transition intensities are assumed to be constant functions of time. There are applications, where the hypothesis of homogeneity may be unrealistic, and a non-homogeneous model is required [20].

2.3. Clustering

Clustering is an unsupervised learning problem, which aims to segment a heterogeneous dataset into homogeneous clusters. As clusters are usually unknown before, it is different to the supervised learning problem classification. The data clustering problem has been addressed by different researchers in many contexts and is an important task in exploratory data analysis [21, 22, 23, 24]. To assign data points to a cluster, it is important to measure their distance to other data points. Different similarity measures, like the Euclidean distance, are used in the literature [25, 26]. Different clusters are formed so the distances between data points in the same cluster are minimal and distances between data points of different clusters are maximal. Clustering sales data was done by [27, 28, 29] where they tried to segment customers with similar buying behaviour into different groups in order to forecast their sales. All of the approaches are using data where every data point already has a selling history and is known to the seller. In this case not all new customers are known to the brand so the challenge lies in determining if they will buy in the future without having a history of purchasing. For that reason, clustering comes into place in combination with a Markov model that adapts the transition probabilities based on the cluster individuals are in. Clustering sales forecasting literature makes use of cluster algorithms such as k-means [30, 28]. As the main goal in this work is to cluster individuals by their buying behaviour the clustering was done in a new way based on the prediction of the individuals buying prediction. In comparison to related research that focuses on existing customers for one sales outlet [27, 28], the approach in this work focuses on forecasting the demand globally, broken down per model on a monthly basis.

2.4. ARIMA Models

To maximize the available data, the new model includes seasonal change in sales, which can be forecasted by a time series of sales conversion over month. This forecast is accomplished using auto regressive integrated moving average ARIMA models. The ARIMA model originated from the auto regressive moving average models. Auto regressive refers to the use of past values in the regression equation for the series; moving average specifies the error of the model as a linear combination of error terms that occurred at various times in the past [31]. An ARIMA model is described by its values (p, d, q), where p and q are integers referring to the order of the auto

regressive and moving average models and d isan integer that refers to the order of differencing [32]. The equation for an $ARIMA(1, 1, 1)$ model is given by [31]

$$(1 - \phi_1 B)(1 - B)Y_t = (1 - \theta_1 B)\,\epsilon_t$$

Where φ_1 is the first order auto regressive coefficient and B is a backward shift operator given by $BY_t = Y_{(t-1)}$ The time series at time t is Y_t, θ_1 is the first order moving average coefficient and ε_t is the random noise at time t [33]. Requirements for the ARIMA model is, that the time series is stationary and there is no missing data within the time series. In the ARIMA analysis, an identified underlying process is generated based on observations of a time series to create an accurate model that precisely illustrates the process-generating mechanism [34]. An extension of this model is the seasonal auto regressive integrated moving average model (SARIMA), which relies on seasonal lags and differences to fit the seasonal pattern [35] and is used in the following dynamic cluster based Markov (DCBM) model.

2.5. Artificial Neural Networks

The new model introduced in the next section includes a new way of clustering data, where an artificial neural network classification is used. For this reason, a short introduction to artificial neural networks is given in the following section.

One of the most commonly used methods in ML is artificial neural networks (ANNs), which try to mimic the biological brain [36]. The equation for each layer of a simple neural network, the multilayer perceptron, is given by [36]

$$y = \varphi(w^T x + b).$$

where w is vector of weights, x denotes the input vector, b the bias and φ is a non-linear activation function. An ANN consists of several connected nodes, called neurons, which receive input from other neurons and send their output to the next neurons. The larger the network, the more input every neuron receives and the more neurons in the next layer receive their output [36]. An important feature of ANNs are that they are nonlinear models as well as universal approximators that provide competitive results by using effective training algorithms. Different training algorithms were used and developed over time: from back-propagation by [37] to newer methods that aim to accelerate the convergence of the algorithm. Although ANNs do not need any prior assumption to build models, as a model is mainly determined by the characteristics of the data, the architecture of the network needs to be predefined [38]. Over time, many suitable extensions have been developed, especially for time series forecasting, such as recurrent neural networks, which are designed to learn time varying patterns by using feedback loops [39].

2.6. Discussion

Thus far, research has not applied multi-stage models to forecast demand for unknown individuals in the database, possibly because the process of sales is usually shorter than the prediction time frame, and thus, this method cannot be used. Depending on the type of product, individuals can be included in the data months before they actually buy. This data can be used to model their process through the various stages to forecast the demand. For products like real estate, boats, or luxury goods the buying process can take weeks to several months. During this

time, the collected data can be used for forecasting. The new DCBM approach can be used to predict the sales pipeline for the future, and it adds time of conversion to the predictions cluster. This added information can be used to improve the forecasting. With the prediction it is possible for a company to break the forecast down on a model as well as regional specific level which supports short term sales and marketing activities in order to sell more products. Furthermore, the proposed approach scores sales opportunities in different clusters, which can be used to influence the conversion by incorporating the resulting clusters into the marketing activities of the business. The methodology of the DCBM model will be introduced in the next section and the case study of a real car manufacturers sales pipeline is described in Section 4, followed by the results and discussion in Section 5.

3. METHODOLOGY

A possible solution to the problem of applying a Markov process onto new unknown individuals, is to compare new unknown customers with existing customers who share similarities. When information can be gained about the individual during the buying process, this information can be used to forecast, for example, buying or building a house or selecting features of a car or a boat. Here, the potential customer is usually known in the customer relationship management system before he or she actually buys. All the information retrieved in this process could be used to forecast on an individual level if the individual will buy or not and thereby support the demand forecasting on an individual level.

Thus far, sales and demand forecasting approaches have been mainly based on time series forecasting approaches, like ARIMA models, if only historic sales numbers are available. Machine learning was used to predict the outcome of the sales pipeline in the past but without an estimated time frame of when the sales opportunities convert [40]. In their research, [40] assumed that the opportunities within the pipeline are more likely to convert by the end of the quarter following targets from the business and only predicted the conversion for a two-week future period using a profile-specific two-dimensional Hawkes processes model.

In survival analysis, an occurrence of an event can be seen as a transition from one state to another, for example, from enquiry to retailed in a sales funnel. Within this research, MSMs are used to forecast sales out of the manufacturers sales pipeline. The different stages individuals go through in their buying process can be modelled as described in Section 2. The new DCBM is suggested: instead of modelling a covariate Z for every individual, the approach here is to cluster the new incoming opportunities in the sales pipeline, described further in Stage 3 of the DCBM Algorithm. This approach was developed because many individuals in the pipeline have no sales, or even stage, history but their features can be used to cluster them according to features from people who already went through the pipeline. To better explain the final model within a business environment and to reduce the curse of dimensionality, a problem occurring by too many covariates, the DCBM is proposed [41]. Each cluster is then modelled with a non-time homogenous Markov model to forecast the number of opportunities which convert over different time frames. The time of conversion information can be gained by calculating the transition matrix for different time frames. For a monthly forecast, the transition matrix for every cluster is broken down on a monthly basis for the next months, depending on the chosen time frame.

All clustering approaches mentioned in Section 2 require the number of clusters to be pre-defined. This requirement leads to the question of how many clusters are suitable for the proposed

approach. Since this question can be a whole research topic itself, it will not be discussed in full detail. Clustering the data based on similarity is possible with approaches like k-means, mentioned in Section 2. As the purpose of every transition matrix is to represent similar buying probability, the clusters should be related to customer groups, which are similar in their buying behaviour. Therefore, the main difference between every cluster should be the sale probability instead of pure similarities in the raw data. Thus, a new approach for dividing data into different clusters is proposed, which starts by estimating the sales probability, which is described further in Stage 2 of the DCBM Algorithm. Afterwards the data is clustered by their sales probability. This approach has the same problem with number of clusters, compared to other clustering approaches, but can improve the prediction of the Markov-based transition model.

Another approach to splitting the data based on sales probability is binary classification, which naturally results in two clusters. To improve the two cluster system and the accuracy in a sales forecast, it is possible to split the first cluster that is likely to sell by using a forecast of the conversion number out of opportunities and split it into two clusters, which is described further in Stage 1 of the DCBM Algorithm. The motivation for the second split is that most sales teams have targets by month, quarter, and year which drive their sales, and there is no additional bonus for selling more than the target, sales teams tend to move potential sales to the next month, quarter, or year if they are close to or have fulfilled their targets [42]. Splitting the first cluster supports the seasonal aspect of conversion and adds that information to the forecast. Therefore, the first group is split into two clusters. One is the size of the prediction from the historic conversion, and the other is defined by the rest of the first cluster, potential sales that are moved to the future to fulfil next month's target or are not treated at all. Figure 1 describes the three clusters and their relation to car sales per month. The blue line represents opportunities created by month as well as their conversion to a sold car on a rolling three months basis after the opportunity was created shown in yellow. The orange line represents the classification boundary where everything above it represents a total number of opportunities that are likely to not convert and the number of opportunities below the orange line represent a total number of opportunities that are likely to convert to a sale. The new idea here is to split the opportunities below the orange line to create two further groups, divided by the yellow line. Everything below the yellow line represents opportunities that are highly likely to convert even more than the ones between orange and yellow.
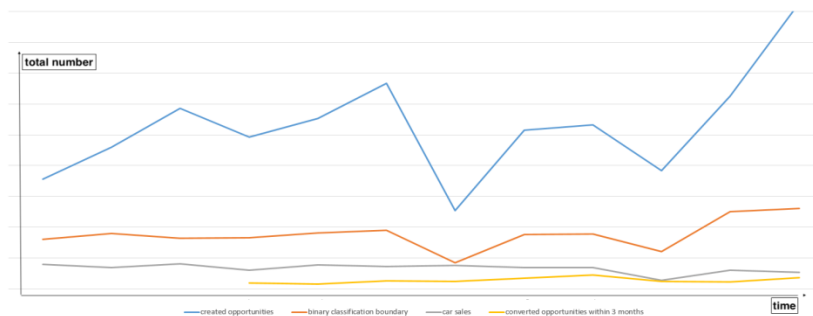
Figure 1: Opportunities, sales and conversions for cluster boundaries

The size of the first cluster is defined by a time series forecast of historic opportunity conversion within a pre-defined time frame of three months as this is the chosen forecasting time span. The number of historic conversions can be forecasted by classical time series approaches, like seasonal ARIMA models, which also contain information about seasonal driven sales by targets.

The calculations between the above explanation is given in the following DCBM Algorithm consisting of three different stages as shown in Figure 2. This diagram shows the data used for training and the flow of information passing between each stage.
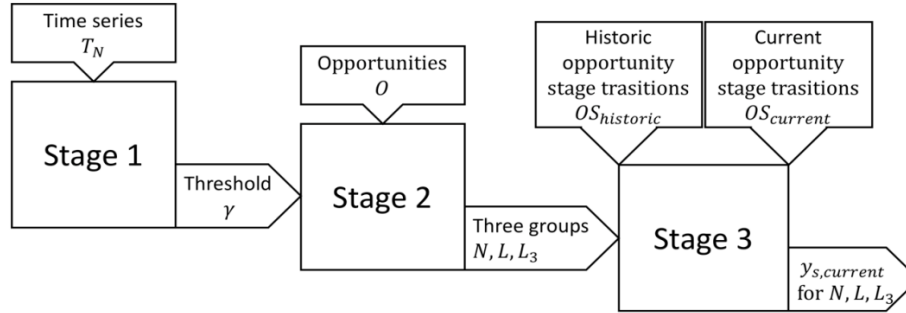

Figure 2: Opportunity stage transitions

Stage 1 describes a seasonal ARIMA model used for forecasting the future conversion of created opportunities per month within the next three months. The model selection is based on Akaike-Information-Criterion of parameters between 0 and 2. The output of Stage 1 is a forecast of opportunities that will converted within three months for the next month. This output will be used in Stage 2 to split the opportunities predicted to convert into two separate groups.

DCBM Algorithm Stage 1: SARIMA opportunity conversion forecast	
Input	Time series T_n=|historic sales conversions out of opportunities per month (2015-today)
Create	Construct $SARIMA\,(p,d,q)(P,D,Q)m$
Initialise	Set SARIMA parameters based on grid search of AIC result between 0 and 2
Algorithm	Time series forecast for following three months
Output	Threshold γ

Stage 2 of DCBM describes a neural network-based clustering which gives every opportunity a probability to convert. Opportunities with a probability over 0.5 are predicted to convert and opportunities below 0.5 are predicted not to convert. The first group of the new DCBM model is defined by opportunities with a probability up to 0.5. The new DCBM approach groups the group likely to convert another time with the output of Stage 1. Ordering the opportunities by their probability with the highest first, the third group consists of the top γ opportunities and the second group consists of opportunities greater than or equal to 0.5 and probabilities lower than from the third group. All three groups are used in Stage 3 of the DCBM Algorithm as for every single group a Markov transition model is applied.

All three groups from DCBM Algorithm Stage 2 are now used to create a Markov transition model for all various stages of the opportunities. Stage 1 and 2 are applied on opportunities older then 12 months as they will not convert anymore in order to get $OS_{historic}$ and on all new incoming opportunities to get $OS_{current}$. The final goal here is calculating the outcome of $y_{s,current}$, representing in which stages the opportunities will be in the next three months.

DCBM Algorithm Stage 2: Neural network-based opportunity clustering

Input	Opportunities $O = (x_1, y_1), ..., (x_n, y_n)$
Create	Build neural network with defined number of inputs, hidden nodes, and outputs
Initialise	Randomly initialize weights for NN
Algorithm	Use gradient descent algorithm to optimize ω_{MP} Obtain predictions $y = f(x; \omega_{MP})$ Cluster Opportunities O into three groups based on predictions N (not likely to convert): $0 \le f(x; \omega_{MP}) < 0.5$ L (likely to convert in more then three months): $0.5 \le f(x; \omega_{MP}) < \gamma$ L_3 (likely to convert within three months): $\gamma \le f(x; \omega_{MP}) < 1$
Output	N, L, L_3

DCBM Algorithm Stage 3: Time homogeneous Markov transitioning model

Input	Historic monthly opportunity stage transitions $OS_{historic} = (s_1, s_2, s_n, y)$ Current month opportunity stage transitions $OS_{current} = (s_1, s_2, s_n)$
Initialize	Construct historic transition matrix $\emptyset O_S$ $\emptyset O_S = \begin{pmatrix} s_1 s_2 & s_1 s_3 & s_1 y \\ . & s_2 s_3 & s_2 y \\ . & . & s_3 y \end{pmatrix}$
Algorithm	For all three groups N, L, L_3 from DCBM Algorithm Stage 2: $y_s = OS_{current} * \emptyset O_S$
Output	$y_{s,current}$ for N, L, L_3

If the number of opportunities classified as a sale in the binary classification is lower than the forecasted number of conversion, the target of the next period likely cannot be achieved from pure demand.

As there is a difference between the demand and actual sales, one can use the demand information to improve the sales forecast. An ARIMA artificial neural network (ANN) [32] can be used to model the residual errors of the ARIMA model with the demand information given above. This solution can also address the potential problem of information asymmetry between dealer and manufacturer, as there can be delayed or missing information within the data. The transition from demand to sales forecast was not applied in the proposed approach but will be researched in the future. The result of the different approaches is compared in the next section using a real dataset.

4. CASE STUDY

The dataset obtained for this study was provided by a luxury car manufacturer in the UK. The main focus of this case study is to predict the opportunity conversions in the automotive industry and thereby improving the sales and demand forecasting on a regional as well as model specific level. This gives the sales and marketing teams worldwide a chance to steer sales by using marketing tools to reach their sales targets. This company's cars are sold exclusively through a dealer network of over 100 dealers worldwide. To sell a car, a dealer needs to create an opportunity in the system, either through converting a lead or manually creating a new

opportunity. Therefore, before selling a car, there needs to be an opportunity in the system. If the dealer only creates the opportunity to directly sell the car or the individual walks into the showroom and directly orders the car, the conversion time from an opportunity to an order is under one month. About 40% of opportunities are created within the same month of the orders. In this case, using opportunities to model an individual transition is difficult as they are created within the last stage. Because every opportunity is created by a human, the quality of the opportunity is much better than online generated leads, which are not pre-qualified. However, if the manufacturer creates opportunities as intended, then it is possible to model the transition between the stages from enquiry (En), qualification (Qu), configure (Co), test drive (TD), negotiate (Ne), commitment (Co), ordered (Or) to closed (Close), which can either be a sold car or a lost opportunity and can transit from stage to stage. Figure 3 illustrates this concept but does not depict all possible transitions, only including the transitions from the first enquiry stage and all possible stage transitions into the last (Close) stage for an easier understanding.

Figure 3: Opportunity stage transitions

Based on all possible transitions, the stages proceed as illustrated in the following transitions matrix in Table 1 between opportunity stages created in two consecutive years in which every value above a diagonal between the top left corner and the bottom right corner displays transitions forward and every value below transitions backwards. The darker the cell, the more transitions happened from and to that stage. The company has sold cars for over a decade, but the stages of the opportunities changed over time as a new CRM system was introduced. Thus, for this study, only data from the two consecutive years onward is used. Data used for training was gathered on a monthly basis for 24 months of total data. In total, the forward transitions for the two years add up to over 100,000, whereas the transitions backward, under the diagonal, total around 3,500. These transitions and those from and to the same stage occurred because of a problem within the system that has since been fixed. Thus, they were not included in the training data as they would introduce bias.

Table 1: Density matrix of transitions between opportunity stages

This matrix is not time dependent, and the transitions from and to the same stage reflect a problem within the system, which has since been fixed. The stages moving backwards were excluded because the system rules prevent them from occurring. Therefore, these transitions were not included. As the CRM system changed over time, over 62,000 opportunities could be used out of approximately 170,000 created so far. This reduction in usable opportunities is due to the changes within the CRM system occurring before the time of the two consecutive years. Also, around one percent of the opportunities are after-sales related and are not used as they are not related to new car sales. Approximately five percent of opportunities could not be used due to a problem within the system as every opportunity should go from one stage to the next stage and not back to the stage at which the opportunity originated. This mistake happened for around five percent of the opportunities, because users sometimes used the system incorrectly and no rule within the system blocked opportunities moving backward. With new rule implementations within the system, this error can no longer happen. Therefore, the opportunities moving backward were excluded. Of the 62,000 opportunities, around 13,000 converted to a sale. The first three months of the following year were used as a test set, and the results were compared on a monthly basis.

Giving the transitions from Table 1 a temporal aspect for the transitions, it takes on average 57 days to change from one stage to another stage. As many opportunities are located in a stage but are in reality lost, every opportunity that did not retail or was cancelled is assumed to be lost after six months. The goal of modelling the transitions is to improve forecasting on a short-term basis, within the next three months, on both a regional and a product level.

The dataset consisted of the features displayed in Table 2 for all opportunities, which were pre-processed as described in the following section before being used in the DCBM algorithm. The categorical variables Region, Model and Lead Source were pre-processed by using one hot encoding, where the categorical variable is replaced with a new binary variable added for each unique categorical value. Related Contact Type was converted to either a 0 for individual or a 1 for already existing customers. Description, Primary Campaign Source, and Secondary Campaign Source were converted to a 0 if they were blank and a 1 if they were filled in. As Sale Probability had missing values, these values were replaced with either the average of probabilities for retailed opportunities or the average probability of lost opportunities. This approach makes best use of Sale Probability compared to leaving it out completely or replacing it with zero. Stage was converted from a categorical variable to a numerical variable between 0 and 1 with similar distributed steps, starting from enquiry and ending with sold with a value of (1) as this is the natural ordering of the stages. An additional feature, Days known, was created to replace Created Date and Contact Created Date, which is calculated by using the difference in days between the day the opportunity was created and the day the contact behind the opportunity was created. This new feature was created in order to see whether the person behind the opportunity was known to the business before and, if so, for how long.

Table 2: Feature overview

Feature	Example	Data type	Missing data
Region	Europe	Categorical	0%
Related Contact Type	Customer	Categorical	0%
Contact Created Date	03/10/2017	Numerical	0%
Created Date	03/10/2018	Numerical	0%
Days known	365	Numerical	0%
Model	Range	Categorical	0%
Stage	Test Drive	Categorical	0%
Lead Source	Website.com	Categorical	36%
Last Stage Changed Date	03/10/2018	Numerical	0%
Description	Wants to buy Y	Categorical	0%
Primary Campaign Source	2018 ABC Campaign	Categorical	31%
Secondary Campaign Source	Launch Campaign	Categorical	78%
Sale Probability	50%	Numerical	32%

5. RESULTS AND DISCUSSION

The prediction aims to discover how many cars out of the opportunities will sell on a monthly basis in the following three months. This prediction can support the sales team on a short-term basis and reflects a live view of sales on a regional as well as model level. Given that information, the sales team can allocate their budget to the regions and models that struggle with future sales, helping them to achieve their targets.

Table 3 compares the results of the proposed approach with a normal Markov transition model throughout all stages. The numbers in the table represent sold cars out of opportunities for the months of January, February, and March and the respective conversion in the first, second, and third months after the opportunity was entered in the system. As the total number of conversions is more important from a business point of view, only this total over the following three months is compared here; however, the comparison for each month indicates advantages in some cases over a traditional Markov model. Comparing the total conversions for the following three months, the proposed approach is closer to the real conversions than the traditional Markov model. In Table 3 the error is calculated by $error = \frac{prediction - real\ conversion}{real\ conversion}$.

Table 3: Opportunity conversion prediction comparison

	Real conversions			DCBM			Traditional Markov		
	Month 1	Month 2	Month 3	Month 1	Month 2	Month3	Month 1	Month 2	Month 3
1st month	109	63	51	92	80	139	208	201	271
2nd month	31	15	31	22	25	63	28	26	26
3rd month	9	20	30	31	39	31	5	16	15
Total	149	98	112	145	144	233	241	243	312
Error				-3%	47%	108%	65%	148%	179%

A drawback of the DCBM is that the time conversion is not completed at the opportunity level. Because it is completed on a cluster level, the approach cannot track when each opportunity will convert. However, the DCBM supports the regions and models potentially at risk and can be supported by marketing activities. Clustering the pipeline as proposed can also be beneficial for

marketing purposes, especially the proposed model that splits the binary classified opportunities based on a time series forecast. For example, one could immediately approach the second stage of the cluster and wait on the first cluster for two months if the opportunities did not convert by then. As the first cluster has the highest probability to convert to a sale, no immediate additional action is required since the customer is predicted to buy. In contrast, in the second cluster, while the probability for a sale is still over the cut off of 0.5, the realistic chances for a sale are lower and could benefit from marketing activities toward the individuals to increase the chance of a sale. The effect of such clusters for marketing could be a future research question related to the marketing department.

On a field trip to different countries, the live sales data was analysed with the help of different dealers in the countries, which supported previous assumptions made from the data and provided new insights into how the companies data is collected. The three clusters were found useful within this framework. Splitting opportunities into three groups was appreciated by sales personnel as this approach made targets and promotions easier to understand and manage on a monthly basis. The findings demonstrate, in theory as well as to the business, that the forecasting was improved and support the company's planning for upcoming months.

As the proposed approach is using opportunities it does not cover all sales as some of them happen before the opportunity is created in the system. To cover this a new approach offorecasting a total sales number will be proposed in the future, where a SARIMA-ANN is used. This approach was chosen as it is more straight forward from a technical point of view to first forecast the opportunities and afterwards use them to forecast overall sales numbers.

6. CONCLUSION

For a company selling its products globally, gaining accurate live forecasts from all countries for all different products is difficult, but it is important to reach its business targets. Using sales pipeline data, which includes stage transitions reflecting the buying process, the DCBM performed better than a common Markov transition approach. Within this approach, a new way of clustering sales data was proposed, which creates three clusters, each differentiated by the customers likelihood to buy a product. The novelty of this approach is that the boundary between the first and second cluster varies based on a time series forecast of conversion, which includes seasonal variations in buying behaviour present in most sales businesses as of given sales targets from the top of the business.

REFERENCES

[1] A. Garnham, Artificial Intelligence: An Introduction. Routledge, 2017.

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, pp. 484–489, Jan. 2016.

[3] R. Michalski, J. Carbonell, and T. Mitchell, Machine Learning: An Artificial Intelligence Approach. Springer, 2013.

[4] U. Franke, D. Gavrila, S. Grzig, F. Lindner, F. Paetzold, and C. Whler, "Autonomous driving goes downtown.," IEEE Intelligent Systems, vol. 13, pp. 40–48, 11 1998.

[5]   J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Zico Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, "Towards fully autonomous driving: Systems and algorithms," in 2011 IEEE Intelligent Vehicles Symposium (IV), pp. 163 – 168, 07 2011.

[6]   C. Chen, A. Seff, A. L. Kornhauser, and J. Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," CoRR, vol. abs/1505.00256, 2015.

[7]   T. Stock and G. Seliger, "Opportunities of sustainable manufacturing in industry 4.0," Procedia CIRP, vol. 40, pp. 536 – 541, 2016. 13th Global Conference on Sustainable Manufacturing Decoupling Growth from Resource Use.

[8]   D. Wright, G. Capon, R. Pag, J. Quiroga, A. Taseen, and F. Tomasini, "Evaluation of forecasting methods for decision support," International Journal of Forecasting, vol. 2, no. 2, pp. 139 – 152, 1986.

[9]   J. Armstrong, Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer, 2001.

[10]   P. Langley and H. A. Simon, "Applications of machine learning and rule induction," Commun. ACM, vol. 38, pp. 54–64, Nov. 1995.

[11]   D. Cox and N. Reid, Analysis of Survival Data. Routledge, 1984.

[12]   E. T. Lee and J. W. Wang, Statistical Methods for Survival Data Analysis. J. Wiley,1971.

[13]   P. Andersen and N. Keiding, "Multi-state models for event history analysis," Statistical Methods in Medical Research, 2002.

[14]   D. Collett, Modelling Survival Data in Medical Research. Chapman and Hall, 2014.

[15]   J. Sullivan and W. Woodall, "A comparison of fuzzy forecasting and markovmodeling," Fuzzy Sets and Systems, vol. 64, pp. 279–293, 06 1994.

[16]   D.-M. Zhu and W.-K. Ching, "A new estimation method for multivariate markov chain model with application in demand predictions," in Proceedings of the 2010 Third International Conference on Business Intelligence and Financial Engineering, BIFE '10, (Washington, DC, USA), pp. 126–130, IEEE Computer Society, 2010.

[17]   P. Hougaard, Analysis of Multivariate Survival Data. Springer-Verlag New York, 2000.

[18]   P. Schweitzer, "Stochastic models, an algorithmic approach, by henk c. tijms (chichester: Wiley, 1994), 375 pages, paperback.," Probability in the Engineering and Informational Sciences, vol. 10, no. 3, p. 463464, 1996.

[19]   A. Kolmogoroff, "Über die analytischen methoden in der wahrscheinlichkeitsrechnung," Mathematische Annalen, 1931.

[20]   C. Gardiner, StochasticMethods. Springer-Verlag Berlin Heidelberg, 2009.

[21]   J. A. Hartigan, Clustering Algorithms. New York, NY, USA: John Wiley & Sons, Inc., 99th ed., 1975.

[22]   H. Spath, Cluster Analysis Algorithms. Ellis Horwood, 1980.

[23]   A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," ACM Comput.Surv., vol. 31, pp. 264–323, Sept. 1999.

[24]   P. Berkhin, A Survey of Clustering Data Mining Techniques, pp. 25–71. SpringerBerlin Heidelberg, 2006.

[25]   P.-E. Danielsson, "Euclidean distance mapping," Computer Graphics and Image Processing, vol. 14, no. 3, pp. 227 – 248, 1980.

[26]   S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," International Journal of Mathematical models and Methods in Applied Sciences, vol. 1, no. 4, pp. 300–307, 2007.

[27]   T. W. Liao, "Clustering of time series data survey," Pattern Recognition, vol. 38, no. 11, pp. 1857 – 1874, 2005.

[28]   P. W. Murray, B. Agard, and M. A. Barajas, "Forecasting supply chain demand by clustering customers," IFAC-PapersOnLine, vol. 48, no. 3, pp. 1834 – 1839, 2015. 15th IFAC Symposium onInformation Control Problems inManufacturing.

[29]   B. P. Kumar, "Improving inventory performance with clustering based demand forecasts," Journal of Modelling in Management, vol. 7, pp. 23–37, 2019/08/16 2012.

[30] P. E. Puspita, T. nkaya, and M. Akansel, "Clustering-based sales forecasting in a forklift distributor," International Journal of Engineering Research and Development, vol. 11, pp. 25 – 40, 2019.

[31] S. L. Ho, M. Xie, and T. N. Goh, "A comparative study of neural network and box-jenkinsarimamodeling in time series prediction," Comput. Ind. Eng., vol. 42, pp. 371–375, June 2002.

[32] P. Zhang, "Time series forecasting using a hybrid arima and neural network model," Neurocomputing, vol. 50, pp. 159–175, 2003.

[33] N. S. Arunraj and D. Ahrens, "A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting," International Journal of Production Economics, vol. 170, no. PA, pp. 321–335, 2015.

[34] G. E. P. Box and G. M. Jenkins, Time series analysis, forecasting and control. Holden- Day, 1976.

[35] R. Yaffee and M. McGee, Introduction to time series analysis and forecasting. Academic Press, 2009.

[36] C. M. Bishop, Neural Networks for Pattern Recognition. New York, NY, USA: Oxford University Press, Inc., 1995.

[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back propagating errors," Nature, vol. 323, pp. 533–536, 10 1986.

[38] P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series,"European Journal of Operational Research, vol. 160, pp. 501–514, 02 2005.

[39] L. Fausett, ed., Fundamentals of Neural Networks: Architectures, Algorithms, andApplications. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1994.

[40] J. Yan, C. Zhang, H. Zha, M. Gong, C. Sun, J. Huang, S. Chu, and X. Yang, "On machine learning towards predictive sales pipeline analytics," in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15, pp. 1945–1951, AAAI Press, 2015.

[41] L. Meira-Machado, J. de Uña-Alvarez, C. Cadarso-Suárez, and P. Andersen, "Multi- state models for the analysis of time-to-event data," Statistical Methods in Medical Research, vol. 18, no. 2, pp. 195–222, 2009.

[42] J. Bouwens and P. Kroos, "Target ratcheting and effort reduction," Journal of Accounting and Economics, vol. 51, pp. 171-185.