Acta Univ. Sapientiae, Informatica **12,** 1 (2020) 70–83

DOI: 10.2478/ausi-2020-0005

Opportunity activity sequence investigations in B2B CRM systems

Doru ROTOVEI
University of West
Timisoara, Romania
email: doru.rotovei80@e-uvt.ro

Abstract. Closing a deal in a business to business environment implies a series of orchestrated actions that the sales representatives are taking to take a prospective buyer from first contact to a closed sale. The actions, such as meetings, emails, phone calls happen in succession and in different points in time relative to the first interaction.

Time-series are ordered sequences of discrete-time data. In this work, we are examining the relationship between the actions as time series and the final win outcome for each deal. To assess whether the behavior of the salespeople have a direct influence on the final outcome of the current deal, we used histogram analysis, dynamic time warping and string edit distance on a real-world Customer Relationship Management System data set. The results are discussed and included in this paper.

1 Introduction

The sales process in the Business to Business (B2B) environment consists of a series of steps and actions intended to take a prospective buyer from initial interest to a closed sale. The series of actions the salespeople take can damage or improve the odds of successfully closing the deal. Therefore analyzing these

Computing Classification System 1998: G.3
Mathematics Subject Classification 2010: 68R15
Key words and phrases: time series; CRM; B2B; machine learning; string edit distance, dynamic time warping;

series of actions can reveal insights into how to successfully close a deal that can be taught and shared with the entire sales team.

Closing a deal in B2B environment is different than selling Business to Consumer (B2C). In B2C, the target is presumably millions of people versus B2B where the potential customers are considerably few. In both B2C and B2B product knowledge is needed however in B2B the product knowledge of the sales representative has to go deeper into design details, advantages, disadvantages and competitors' knowledge as the buyers understand the complexity of the products and their sophistication. In B2C usually one decision maker is involved, in B2B orders are considerably higher in monetary value, multiple decision-makers need to be convinced which leads to a longer time period to close the deal [3, 21]. Therefore a system is needed to keep track of all these activities. The tracking of all the information, requirements, products of interest, notes and, in general, entire interaction with the potential customer is tracked using an opportunity entity inside a Customer Relationship Management (CRM) System [8].

Each activity (such as meetings, emails, phone calls) that a sales representative is performing for an opportunity is usually recorded with their associated notes for historical and recollection reasons. We can note that each activity is performed in a precise point in time relative to the creation of the CRM opportunity. Furthermore, the time gap between activities and the type of activity (email, phone call) could play a role in the final outcome of the deal.

How much information is recorded in the series of activities for each opportunity? Is there a pattern of a successful sale in the way a deal is won? Can we distinguish the series of activities that lead to lost sales versus won sale? We set ourselves to find out the answer to these questions by doing a time series analysis of the activities performed for each won and lost opportunity.

Our hypothesis is that won deals and lost deals can be clustered together, in other words, there is a distinctive pattern of activities performed for won deals and for lost deals. This hypothesis, if true, could help us improve the B2B sales prediction models that we previously studied in [17, 16].

This paper is organized as follows: section 2 describes related work on B2B time series analysis, section 3 introduces our methodology and the problem domain while sections 4 and 5 presents our results, conclusion and future work.

2 Literature review

In the last few years, researchers have become increasingly interested in improving B2B selling using CRM data analytics [14]. Therefore in this section, we will look first at the B2B selling and the influencers of a successfully closed deal followed by sales forecasting and the methods used to predict future business income.

2.1 B2B selling

B2B sales are driven by the relationship that is established between the salesperson and the prospect. One study found that the relationship is stronger and the actions more impactful when the product or service is more critical to the customer and when the sales person is directly interacting with an individual decision maker instead of the firm [13]. In another study, the performance of 816 salespeople in 30 sales organizations has been conducted to research the influence of the company strategy towards the final sales outcome. It has been found that the company strategy influences considerably the individual salespeople performance especially related to the prioritization, customer orientation and value-based selling [19].

The clarity of the sales role, no ambiguity and lack of conflict, has a direct positive influence on the job performance which in turn influences the quota attainment and the individual opportunity success rate [9].

The ethical behavior of salespeople has been studied in [20] where it has been found that ethical behavior is linked to increase sales and customer satisfaction as well as product or service quality. Likewise, the sales person adaptability during the sale has a strong correlation to the success of the sale [2]. Adaptability requires the examination of the internal and the external activity that trigger changes in behavior to successfully close a sale.

There is also momentum in a closing deal. After each stage and after each customer interaction a successful sales representative will ask for an agreement for the next step. If the potential customer is moving along and the set milestones are hit, more than likely the successful sale will happen [4].

In this study, we are looking at the behavior of salespeople as recorded in the series of activities performed on each opportunity and their relationship to the final outcome. The behavior, even influenced by the company strategy or the adaptability of the sales person, if consistent can shed light on the steps necessary to successfully close a deal.

2.2 Sales forecasting

Forecasting is the process of predicting future events based on the information and events that already occurred [1]. Sales forecasting is a prediction of future sales performance using the information known today from marketing conditions to sales pipeline analysis.

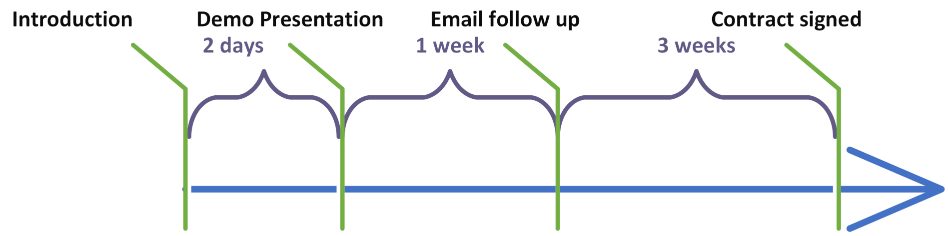

Figure 1: An example of activities succession to close an opportunity

As sales are the lifeline of any business, a significant amount of research has been done to predict future sales using time series analysis in different industries such as fashion, e-commerce, appliances to name a few [12, 11, 22]. Even more, combining forecasting from different models has shown to perform better than a single model forecast for most time series [6].

However, to best of our knowledge, no study has looked specifically at the impact of the individual activities for each opportunity and how the implied behavior might affect the final lost/win outcome.

3 Problem domain and methodology

B2B sales take time to close. To keep track of the progress on each opportunity, sales representatives are recording notes along with their activities. The customer interaction history might include quotes and proposals, product codes and volumes needed, delivery requirements or product configuration information. All in all, the opportunity entity contains the stages, milestones and key activities performed for each opportunity.

Sales representatives, over time, develop a set of steps and activities that they execute to close a deal that could be as personal as a fingerprint. For example one sales representative might use a demo meeting, following with an email, waiting for a week and making a phone call, whereas another sales representative might be more aggressive and after the demo follow up with a phone call and two days later with another voice mail. Part of the behavior of the sales representative is recorded in the succession of the activities performed and also in the length of time between each activity until the opportunity is closed (see Fig. 1).

Our study builds up by starting first with histogram analysis of the lost/won deals, subsequently by time series analysis using dynamic time warping and wraps up with string edit distance.

3.1 Histogram analysis

A histogram shows the frequency distribution of a continuous data set. Using the shape bar bins which represent the intervals of the continuous data, the data is plotted bidimensionally against the frequency [7].

In the first part, we created histograms of the won deals and of the lost deals. The histograms, if different, can show us these differences between the activity patterns for lost deals compared to the activity patterns for the won deals.

3.2 Dynamic time warping

Our subsequent approach to discover similarities between won activities and lost activities was to use dynamic time warping (DTW) which is an algorithm used for comparison of time series. In essence, given two time series of activities the algorithm stretches or compresses the series along the time axis in order to resemble each other as much as possible while simultaneously measuring the similarities between the two time series.

Formally, given two time series of length n and respectively m:

$$\begin{aligned} X &= x_1, x_2, \ldots, x_n \\ Y &= y_1, y_2, \ldots, y_m \end{aligned} \tag{1}$$

we construct a wrap path W

$$W = w_1, w_2, \ldots, w_K \qquad max(m, n) \leq K << m + n \tag{2}$$

where K is the length of the wrap and the k element of the wrap path is:

$$w_k = (i, j) \qquad i = 1 \ldots n, j = 1 \ldots m \tag{3}$$

the distance of the wrap path W is:

$$Dist(W) = \sum_{k=1}^{K} Dist(w_{ki}, w_{kj}) \qquad i = 1 \ldots n, j = 1 \ldots m \tag{4}$$

The optimal warp path is the wrap path with the minimum distance[18].

3.3 String edit distance

Similar to a DNA sequence encoded with the four letters AGTC that undergoes insertions, deletions substitutions and transpositions, we could encode all the activities performed on an opportunity as a succession of letters and subsequently for each succession calculate the distance between any other succession using a string edit distance metric. The main idea is to compare two DNA sequences and see how closely they are to each other.

3.3.1 Damerau–Levenshtein string edit distance

Damerau and Levenshtein studied spelling errors and discovered that almost 80% of the spelling errors are at distance one in the metric that carries their names [5]. The metric is a function from an alphabet combination of characters to an integer value [10]. The created metric evaluates how many operations are needed to transform string s1 into string s2.

The distance d(s1,s2) between two strings can be a combination of the following operations:

- insert a character,
- delete a character,
- substitute a character with another from the same alphabet,
- transposition of two adjacent characters.

Although can be a high number of combination of these operations to convert s1 into s2, the metric returns the length of the shortest sequence as the distance between the two strings. For example, in Fig. 2, to transform the text ABACDEAAB into BACDEAEAB we need two operations a delete operation and an insert operation: **A**BACDEAAB $\Longrightarrow$ BACDEAAB $\Longrightarrow$ BACDEA**E**AB. Therefore $\mathsf{d}(\mathsf{ABACDEAAB}, \mathsf{BACDEAEAB}) = 2$.

We used this metric to calculate the distance between the list of activities performed on won deals and the list of activities performed on lost deals in two flavors. First, we did not consider the time elapsed between activities. On the second approach we encoded any week with no activity with a 0. This way we take into account not only the type of activities and their succession, but also the time component that captures periods of inactivity.

4 Results

4.1 The dataset

The data set [15] used for our research represents B2B sales of an ERP software that is sold globally. The data set has 276 deals, 153 lost and 123 won deals in the period 2009-2016 with 19082 activities.

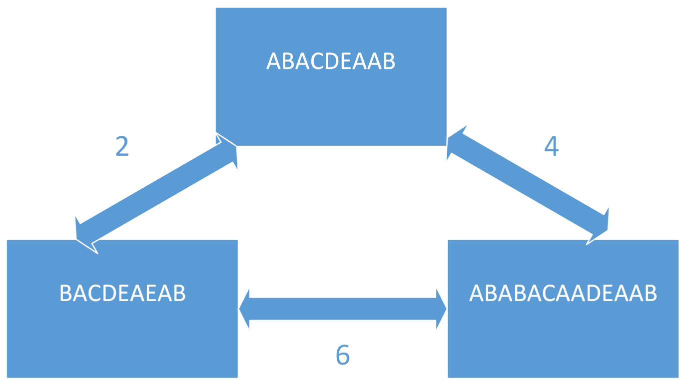

Figure 2: An example of activity succession alignment using Damerau–Levenshtein string edit distance

A summary statistics of the data set is present in Table 1.

Furthermore, the quartiles of the number of days to a closed deal is shown in Table 2.

4.2 Histogram analysis results

We used two approaches to preprocess the opportunities data. The first approach was to extract the minutes between two consecutive notes for each lost and won opportunity.

A second approach for histogram analysis was to use the number of activities the sale representative is performing each week for each opportunity. This approach is meant to reduce the variance due to a too granular look when using the number of minutes between activities. This approach solves also the weekend problem as in general there are no activities in the weekend and national holidays (however, if there are we don't need to treat them specially). Using the week level aggregation could reveal more clear weekly behavioral patterns.

Examining the histogram in Fig. 3 we notice that there are more activities for the lost deals than the won deals (frequency bypasses 5000). However, this is correlated with the fact that there are more lost activities than won activities in the data set. Other than these insights, the histograms show that the wait time between activities is similar for the won deals and lost deals with just a few tiny spikes on the lost deals side. In this context, there are no partic-

Value	Won Deals	Lost Deals
Opportunities	123	153
Activities	12129	6953
Avg(#activities/opportunity)	98	45
Max(#activities/opportunity)	286	153
Min time to a closed deal	4	1
Avg time to a closed deal	47	38
Max time to a closed deal	195	155

Table 1: Data set statistics

Quartile	Won Deals	Lost Deals
0%	4	1
25%	23	23
50%	47	38
75%	98	58
100%	195	155

Table 2: Quartiles of how many days it takes to close a deal

ular idiosyncrasies between the two sets to extract an obvious differentiating behavior. One of the reasons could be that the minutes between activities is a too detailed level of analysis and we might need to zoom out. This prompted us to explore the number of activities per week as well.

The results analysis for the number of activities per week in Fig. 4 shows that in the first week a lot of activity happens for both won and lost deals. However, after the first couple weeks the pattern is slightly different for won and lost deals. The won deals have a constant stream of activities, whereas the lost deals, similar to the minutes between activities histogram, have some spikes.

This prompted us to further examine the patterns of activities for the lost/won deals using dynamic time warping time series analysis and string edit distance to further isolate and measure the possible patterns.

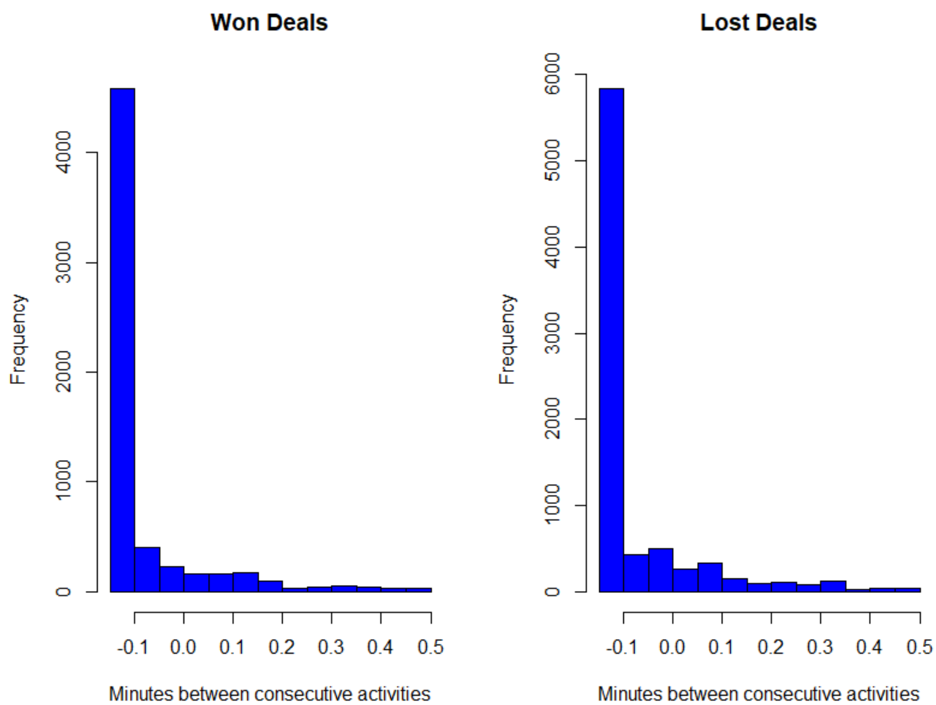

Figure 3: Histogram of minutes between activities for lost/won deals

4.3 Dynamic time warping results

Dynamic time warping is an algorithm used to measure the similarity between two sequences which may vary in speed or time. To prepare the data for DTW exploration we used the same process as above for counting the number of notes per week for each won and lost deal.

We conducted a quantitative analysis by measuring DTW using Euclidean distance and also a variant using standardization. We would have expected the won deals to be close together with a low DTW number and the won versus lost deals to have a higher number. However, as seen in Table 3, the average distance between won deals is 77, the average distance between lost deals is 35 and between lost and won deals is and in between value of 35. This last value we would have expected to be very high which it would prove that we can separate the lost deals from won deals using DTW.

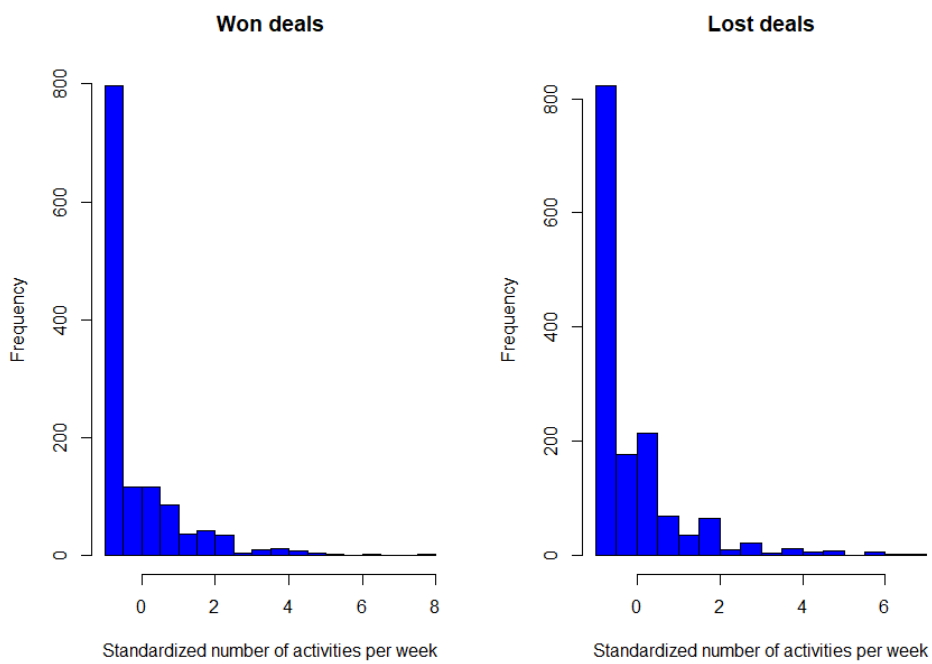

Figure 4: Histogram number of activities per week for won and lost deals

Yet, these findings suggest that our results did not provide convincing evidence towards our hypothesis which means that a pattern cannot be established between the won deals and lost deals. To further prove the new hypothesis, we turned to string edit distance for an additional perspective analysis.

4.4 String edit distance results

To use Damerau–Levenshtein string edit distance we encoded each activity type with a letter from the ASCII code. For our data set we have 28 different activity types. Once we assigned a character to each activity type we create the sequence of activities for all the opportunities lost and won. The process is illustrated in Table 4.

Once all the activity letter sequence was calculated for each opportunity we calculated the distances between won activities, lost activities and between the won and the lost.

Deals	Avg DTW distance	Avg DTW normalized distance
All won deals	77	2.96
All lost deal	35	0.31
All won vs all lost	61	1.74

Table 3: DTW distance between deals

Activity	Encoding	Sequence
Phone call	A	A
Email	B	A**B**
Phone call	A	AB**A**
Customer visit notes	C	ABA**C**
Proposal preparation	D	ABAC**D**
Fax	E	ABACD**E**
Phone call	A	ABACDE**A**
Phone call	A	ABACDEA**A**
Email	B	ABACDEAA**B**

Table 4: Activities encoding sequence

The evidence we were looking for was that the won opportunity activities are clustered together showed by a low distance, the same for lost opportunity activities and in contrast the distance between won opportunity activities and lost opportunity activities is considerably higher. This would mean that we could group the won activities and the lost activities and therefore could establish a pattern for each group of activities.

As we can see from table 5 the average distance between all won deals was 77 and the average distance between all lost deals was 61. These numbers prove that that the succession of activities for the won deals is much more different than for the lost deals. When we calculated the distance between the succession of activities for the won deals against the succession of activities for the lost deals we obtained 42. As can be seen, the numbers match the findings we attained with dynamic time warping: a high distance between the won deals, a low distance for the lost deals and an in-between distance for the won versus lost deals.

The initial approach that we used for string edit distance did not take into account the time element. For example, a phone call followed by an email can

Deals	Avg string edit distance
All won deals	77
All lost deal	61
All won vs all lost	42

Table 5: Damerau–Levenshtein distance between deals

Deals	Avg string edit distance
All won deals	1022
All lost deal	334
All won vs all lost	748

Table 6: Damerau-Levenshtein distance between deals with time series

be done after 1 day or after 1 month. Obviously, it is much better to do it sooner than later. For this reason, we explored a second approach of calculating the string edit distance by encoding weeks with no activities as 0. Therefore a sequence of AB can be A00000B if 5 weeks passed with no activities on the active opportunity. The results of the string edit distance with 0s for weeks with no activity are shown in Table 6.

Although the distances are larger, the results in Table 6 strengthen again the argument that a pattern could not be established. Therefore, contrary to our expectations, our hypothesis that there is a pattern of actions for won deals (and lost deals) does not hold. The general picture emerging from this analysis is that current actions taken cannot predict the future deal outcome.

5 Conclusion and future work

In this paper, we looked into the behavior of the salespeople captured by the successions of activities to determine their influence towards the final outcome. We started with the hypothesis that the pattern of activities for won deals might be different from the patterns of activities for lost deals. However, the results yielded some interesting findings.

The analysis of the succession of activities for lost deals and for the won deals shows that the sales representatives win their deals in very different ways as exhibited by the high distance value for both DTW and Damerau–Levenshtein distance. In contrast, the deals are lost in a more comparable fashion revealed by a low DTW and Damerau–Levenshtein distance.

More surprising however, is the fact that the lost deals patterns are closer to the won deals than the won deals themselves. We can infer from these results that the sales representatives don't behave differently, they behave the same way for the won deals and for the lost deals and their behavior is not correlated with the final outcome i.e. they treat each deal the same putting the same amount of effort to close the deal, not knowing if it will be lost or won which is revealing and encouraging for the sales team.

However, our data set was limited to one B2B CRM instance. Future research will have to clarify whether similar results could be obtained with a different B2B CRM data set. Furthermore, we examined the time series without adding any contextual knowledge (the size of the deal, the period during the quarter, the country of the customer). Adding context might offer additional insights that we plan to address in more detail in a future work.

References

[1] J. S. Armstrong, *Principles of forecasting: a handbook for researchers and practitioners* , Springer Science & Business Media, 2001. ⇒72

[2] C. D. Bodkin, C. P. Schuster, A Preliminary Investigation of the Predilection for Adaptive Behavior and Sales Success, *Proceedings of the 1988 Academy of Marketing Science (AMS) Annual Conference*, 2015, pp. 291–295. ⇒72

[3] E. Bridges, R. Goldsmith, C. F. Hofacker, Attracting and retaining online buyers: comparing B2B and B2C customers, *Advances in electronic marketing* **4772** (2005) 1–27. ⇒71

[4] F. Buttle, *Customer Relationship Management: Concepts and Technology* , Sydney: a Butterworth-Heinemann Title, 2009. ⇒72

[5] F. J. Damerau, A technique for computer detection and correction of spelling errors, *Communications of the ACM* **7,3** (1964) 171–176. ⇒75

[6] M. Gahirwal, Inter time series sales forecasting, *arXiv preprint arXiv:1303.0117* (2013). ⇒73

[7] D. Griffiths , *Head first statistics* , O'Reilly Media, Inc., 2008. ⇒74

[8] K. Jaya, R. Vadlamani, Evolutionary computing applied to customer relationship management: A survey *Engineering Applications of Artificial Intelligence* **56** (2016) 30–59. ⇒71

[9] T. A. Judge, C. J. Thoresen, J. P. Bono, G. K. Patton, The job satisfaction–job performance relationship: A qualitative and quantitative review, *Psychological Bulletin* **127,** 3, (2001), 376–407. ⇒72

[10] V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, *Soviet physics doklady* **10,8** (1966) 707–710. ⇒75

[11] N. Liu, T. Choi, C. Hui, S. Ng, S. Ren, Sales forecasting for fashion retailing service industry: a review *Mathematical Problems in Engineering* **2013**. ⇒73

[12] P. Pai, C. Liu, Predicting Vehicle Sales by Sentiment Analysis of Twitter Data and Stock Market Values *IEEE Access* **6** (2018) 57655–57662. ⇒73

[13] R. Palmatier, R. Dant, K. R. Evans, D. Grewal, Factors influencing the effectiveness of relationship marketing: A meta-analysis *Journal of marketing* **70, 4** (2006) 136–153. ⇒72

[14] A. Rainer, P. Thomas, Successful practices in customer relationship management, *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, Hawaii, USA, 2004. ⇒71

[15] D. Rotovei, B2B ERP sales activities data set, *B2B ERP sales activities data set*. ⇒75

[16] D. Rotovei, V. Negru, *Improving Lost/Won Classification in CRM Systems Using Sentiment Analysis*, *Proc. 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Timisoara, Romania, 2017, pp. 180–187. ⇒71

[17] D. Rotovei, V. Negru, *A methodology for improving complex sales success in CRM Systems*, *INnovations in Intelligent SysTems and Applications (INISTA)*, Gdansk, Poland, 2017, pp. 322–327. ⇒71

[18] S. Salvador, P. Chan, Toward accurate dynamic time warping in linear time and space, *Intelligent Data Analysis* **11,5** (2007) 561–580. ⇒74

[19] H. Terho, A. Eggert, A. Haas, W. Ulaga, How sales strategy translates into performance: The role of salesperson customer orientation and value-based selling *Journal of Industrial Marketing Management* **70, 4** (2006) 136–153. ⇒72

[20] A. S. Tolba, M. Iman, A. M. Hakim, Impact Of Ethical Sales Behavior, Quality and Image on Customer Satisfaction and Loyalty: Evidence From Retail Banking in Egypt, *International Journal of Management and Marketing Research* **8,2** (2015) 1–18. ⇒72

[21] K. Vinod, R. E. Gagandeep, Business to business (b2b) and business to consumer (b2c) management *International Journal of Computers & Technology* **3, 3b** (2012) 447–451. ⇒71

[22] D. Wei, P. Geng, L. Shuaipeng, L. Ying, A prediction study on e-commerce sales based on structure time series model and web search data, *Proceedings of the 126th Chinese Control and Decision Conference (CCDC)*, 2014, pp. 5346–5351. ⇒73

Received: January 3, 2020 • Revised: March 2, 2020