

## PSEUDO-MULTIVARIATE LSTM NEURAL NETWORK APPROACH FOR PURCHASE DAY PREDICTION IN B2B

UDC (004.032.26:(005.52:330.133.1))

Milica Ćirić<sup>1</sup>, Bratislav Predić<sup>2</sup>

<sup>1</sup>University of Niš, Faculty of Civil Engineering and Architecture, Republic of Serbia

<sup>2</sup>University of Niš, Faculty of Electronic Engineering, Republic of Serbia

**Abstract.** *This research focuses on trying to predict the moment of the next purchase for a customer in vendor-customer B2B scenario using an LSTM neural network and comparing prediction results from different input features. In a previous research we performed predictions for a specific customer product pair and used previous purchases for that pair as input data, but the number of such previous purchases was often very limited which resulted in low accuracy of predictions. By aggregating purchase data for all products a customer purchased, we were able to get more precise predictions of the next purchase. Additionally, expanding our input feature set yielded even better results. We performed an evaluation of LSTM networks trained with the most successful combination of input features for a six month period. Each of the networks was trained with purchase data up to the starting point of the selected period and the predictions were performed, after which additional input for the following seven days was added to the network. This process was then repeated for the entire six month period and a slight downward trend can be noticed for error metrics, leading to the conclusion that the network would perform even better over time with the addition of future purchases.*

**Key words:** *Purchase prediction, time series, long short-term memory neural network*

### 1. INTRODUCTION

Businesses that sell products to customers keep track of their inventory at all times. In an ideal scenario, an inventory should always contain just enough of each product, but not too much, especially if that product has a use by date. In order to perform a proper inventory planning, businesses need to have an estimate of customer purchases that will happen in the near future. One approach to making an estimate is predicting future purchases based on previous purchase data. An additional benefit to customer purchase

---

Received November 16, 2020

**Corresponding author:** Milica Ćirić

University of Niš, Faculty of Civil Engineering and Architecture, Aleksandra Medvedeva 14, 18000 Niš, Republic of Serbia

E-mail: milica.ciric@gaf.ni.ac.rs

prediction is the possibility of personalized directed marketing. If we can predict customer behavior, we can offer them special conditions that might affect their buying decision.

Personal purchase decisions are often made impulsively which makes them harder to predict. Purchasing for a business, on the other hand, is a more formalized process, performed rationally when an actual need exists, and influencing the customer is the most likely when that need reaches its peak [1]. Supervised learning predictions based on anonymized and aggregated transactional data from a B2B business network can be used in inventory/procurement planning [2].

In our research we focused on predicting next purchase for a customer in a B2B setting using purchase data from a medical supply vendor. Initial experiments targeted customer-product pairs [3], [4] and made predictions by using input data only for those pairs. Due to a limited supply of customer-product pairs with a significant number of repeated purchases, achieved prediction accuracies were low for both statistical methods used and LSTM neural networks. After analyzing a large quantity of purchase data for customer-product pairs and aggregated purchases for customers, we noticed that the time series of all customer purchases are more regular than those that are for customer-product pairs. Almost all purchases for a customer contain multiple products, but the products are not always the same, presumably due to different needs and inventory state. As a result, the time series that contains all purchases contains more data points, whereas the time series for specific products have larger gaps. This led us in the direction of aggregating customer purchases and focusing on when the customer will make the next purchase, instead of trying to also predict the product which will be ordered. By aggregating purchases for a customer, our training and testing corpuses became much larger and the number of neural networks required dropped significantly. Results of performed experiments show that all accuracy metrics improved greatly and the training time was shortened.

In order to produce even better results, we have introduced some additional input features for the LSTM neural networks. The setup that achieved the highest accuracy in predicting the next purchase day was then tested with a sliding six month period of purchase records. The original data corpus consists of about two and a half years of purchase records for a medical supply vendor. In all of our testing scenarios we used a six month period as test data. This length of the period was chosen in order to make sure that the next purchase was detected. Customers/customer-product pairs that did not have a repeated purchase in that period were discarded from evaluation since our premise was that they either continued purchases with a different supplier or ran out of business. For the sliding testing we started with training the networks with the first 18 months of purchase transactions and tested them with the following 6 months of purchase transactions. Then, we trained the networks with additional input, specifically purchases from the following seven days, and slid the testing period for seven days. We continued with this process until we reached the end of purchase data. The results of sliding the evaluation period enabled us to track accuracy metrics over time and register if there are consistency and/or trends in these metrics.

## 2. RELATED WORK

In order to perform a proper inventory planning businesses need to have an estimate of customer purchases that will happen in the near future. Predicting this behavior based on previous purchases is a topic of many researches [3], [5]. In [5] a sequential time series is created for all relevant customer product combinations. Multiple deep learning models, as

well as tree based models are then evaluated with different hyperparameter combinations and deep learning models outperform tree based models in terms of accuracy.

Customer base is a term that refers to existing customers of a business, especially those that are the most profitable. Businesses try to keep such customers and encourage them to perform additional purchases. [6] compares different approaches to customer base predictions - stochastic methods represented by Pareto/NBD and machine learning represented by artificial neural networks. Evaluated on three datasets, neither approach consistently outperforms the other in all predictions.

When predicting future customer behavior most researches focus on known/existing customers [3], [5], [7], but some also consider anonymous customers which, in some e-commerce systems, make up to 50% of all sessions [8].

In [9] the authors collected tweets that mentioned mobile devices or cameras using the Twitter API and tried to predict which users will buy these devices. LSTM neural network performed best in the customer product prediction, while a feed/forward neural network was the most successful in determining relevance to customer purchase behavior. Utilizing the sequential nature of tweets was proven to be a significant factor in predicting a purchase.

The research presented in [10] deals with a multi-task LSTM neural network for predicting day of the week, time of the day and product category using online grocery shopping transactions as input data. Out of tested network settings and feature combinations, none produced the best results for all three tasks and the product category proved the most difficult to predict.

Predicting next month purchases using machine learning techniques using a customer relevant features set derived from times and values of previous purchase is the focus of the research presented in [11]. The features are updated every month and prediction is performed if the customer will make a purchase within a certain time frame in the near future. The main challenge that emerged in this non-contractual setting is how to differentiate customers who have ended their relationship with the firm from those who are simply in the midst of a pause between transactions.

The authors of [12] identify two main challenges in modeling online purchase behavior: multi-scale temporal patterns which may be inter-correlated and dynamic arbitrary dependencies of categories. As a response they propose a general framework Graph Multi-Scale Pyramid Networks (GMP) for online purchase prediction. Experiments with three datasets show that GMP achieves the best results compared to multiple state-of-the-art approaches in all categories.

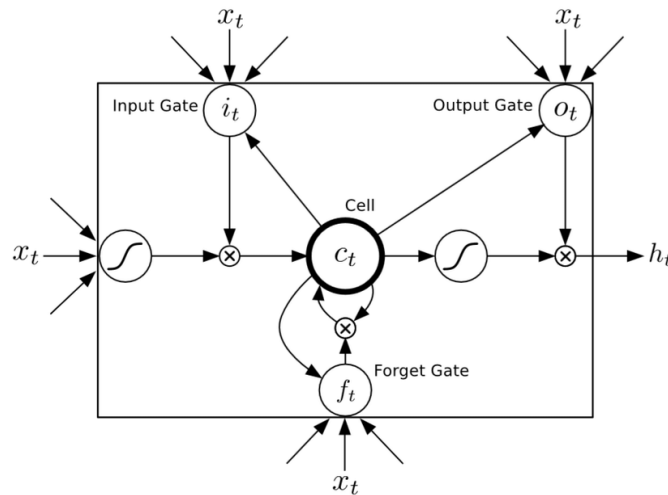
[7] focuses on optimizing Market Basket Analysis (MBA) through data cleaning and mining data association rules but also by feed forward neural network classification. Comparison to other predictive models for MBA shows significant improvement of F-measure using the proposed model.

The authors of [13] identified that the customer purchasing decision could be predicted using a set of 22 hashed features from decontextualized data with great certainty using Naïve Bayes classifier and Random forest. Also, there was a lot more interaction of different features describing the purchasing context interact in a more nuanced way when the customer goes ahead with the purchase.

In [14] the researchers designed a multi-layered back propagation neural network based on the assumption that the behavior of the users looking to buy a product differs from the behavior of those just browsing the store. The input data used consisted of user sessions reconstructed from Web server logs and a high precision rate was achieved in predicting whether the user had made a purchase.

Our chosen approach for purchase prediction is the use of Long Short-Term Neural Networks (LSTM). LSTM neural networks are a type of Recurrent Neural Networks, whose main advantage is the ability to register longer dependencies that exist in input data by solving the Vanishing Gradient Problem.

They were initially proposed in [15] in 1997. Since then, there have been multiple proposed variations, with the most important one being the introduction of the forget gate in [16] and this is the most widely used variation. In contrast to Feed forward Neural Networks, which only pass data forward, Recurrent Neural Networks have returning connections which enable using the old cell state in addition to new cell input. Hence, it can be said that the neural network has some form of memory and output at any given time is based on new and past input. As time steps pass, the influence of older inputs fades, which is why more recent inputs affect output more than older ones. However, RNNs suffer from the Vanishing Gradient problem [17], i.e. due to the multiplying of a large number of small values, the gradient can become a very small value. Vanishing Gradient causes insufficient weight updates in the network, which results in difficulty in learning long-term dependencies [17]. Therefore, RNNs are theoretically able to remember long-term dependencies, but in practice this is not the case. The main advantage of LSTM neural networks is the ability to store long-term dependencies in data [15]. The architecture of a LSTM cell is shown in Fig. 1. Each cell is essentially a memory block with three multiplicative gates: input gate, forget gate and output gate, and they control what part of input, cell state or output will be used in further calculations, and what part will be discarded. The input gate protects the memory content from irrelevant input [18]. The forget gate determines when to forget previous state by resetting it or by gradual fading and the output gate prevents the cell state to perturb the rest of the neural network [18].



**Fig. 1** Long Short-Term Memory Cell [19]

The LSTM memory cell shown in Fig. 1 is described with following equations [19]:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (4)$$

$$h_t = o_t \tanh(c_t) \quad (5)$$

where  $\sigma$  is the sigmoid function,  $i, f, o, c$  and  $x$  are activation vectors for input gate, forget gate, output gate, cell and cell input,  $h$  is the hidden vector,  $b$  the biases and  $W$  the weights matrices. Subscripts for each of the weight matrices denote which connection they apply to.

### 3. PROBLEM SETUP AND PROPOSED SOLUTION

The idea behind this research and some of its previous stages is helping a vendor in a B2B scenario predict purchases in the near future (i.e. next week) and use this information in two main ways. First, predicting future demand can be very helpful in planning the procurement process. Secondly, the knowledge about future purchases can be used in the purpose of advertising directly to the customer in a personalized style. If future purchases can be predicted with a relatively high accuracy, we can also detect those purchases that were supposed to happen, but did not. Since making an order in a B2B scenario is not a matter of “want”, but a matter of “need” [4], the absence of purchases from a customer probably means that they decided to switch the supplier and a swift reaction (i.e. offering better purchasing conditions) from the vendor might make them reconsider. In this scenario the prediction process would be performed once every Sunday and the output, i.e. predictions for the following seven days would be used by the company for some customer-targeted actions. Next Sunday, purchase from the last seven days would be added as input to the neural network and the prediction process would be executed again, producing new predictions. Additionally, predictions from the previous iteration could be analyzed and compared to the actual data. This cycle could be continued as long as the need exists.

According to the aforementioned requirements the prediction process was designed as shown in Fig. 2. Data was collected from a medical supply company, around three million transactions from a two and a half year period. All customer data was anonymized prior to further processing. Each transaction contains, among other things, identifiers for the customer and product purchases, date of the transaction and product quantity. However, although these transactions form time series, they are irregular and need to be transformed before feeding them into a neural network. We chose to transform these transactions into times series consisting of the number of days that pass between two consecutive purchases. In [3] a separate time series was formed for each customer-product pair, resulting in separate neural network for each of the pairs. After analyzing testing results and purchasing transactions, we decided to shift to customer level. Individual time series was formed for each customer, consisting of all purchases they made. Accordingly, one LSTM neural

network was trained for each customer, leading to a much smaller number of networks than in the previous research. As mentioned, we operated with the assumption that the absence of purchases from a customer means they switched the supplier. For evaluation, the testing period was defined as a six month period in order to ensure the inclusion of the next realized purchase. First experiments were performed using the first two years of the purchasing data as training data and the remaining six months as testing data. In later experiments, we emulated a six month use of the designed system by starting 18 months after the first transaction and then adding a week's worth of purchasing transactions for each testing iteration while sliding the testing period.

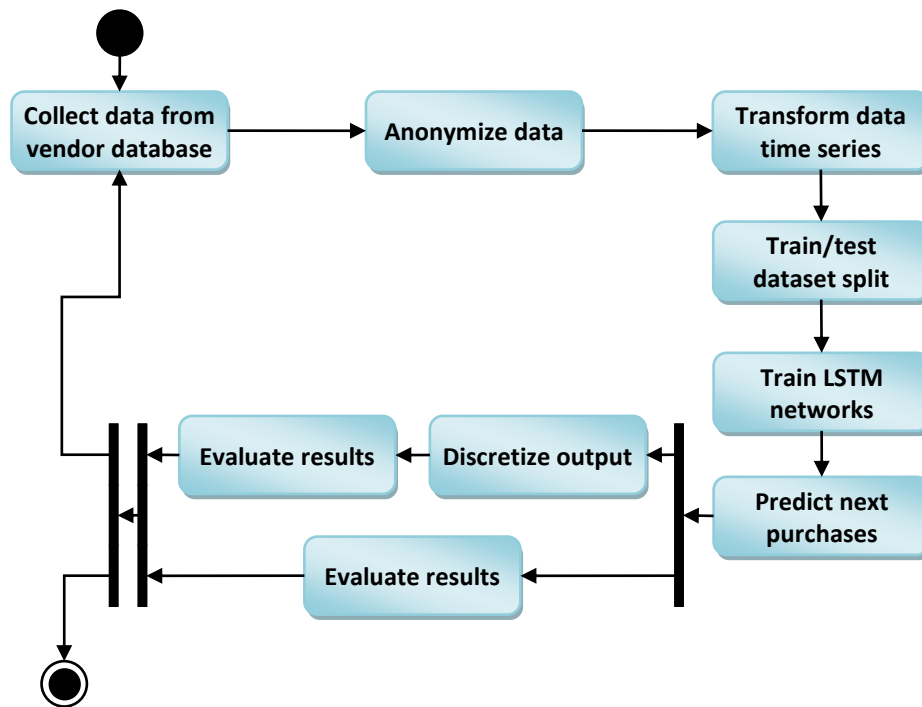


Fig. 2 Activity diagram of the purchase prediction process

After the dataset is split to training/testing portions, one LSTM neural network was trained for each customer and a prediction for the next purchase was generated. Predictions were generated as a number of days that will pass until the next purchase. In agreement with previous researches [3], [4], we discretized predicted values to two classes: “purchase expected” in the following seven days and “purchase not expected” in the following seven days. For these two classes precision and recall were calculated and then compared to earlier results. However, we also evaluated Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) for predicted values (before discretization) for each of the setup.

In a one-off experiment, the prediction process ends after the evaluation step, but in the system’s intended use scenario the process would be repeated each week. We tried to recreate this scenario by emulating a six month use of the system. Fig. 2 shows that in

each iteration all steps are repeated, from data collecting, anonymizing, transformation, dataset split, training to discretizing and evaluation of the results. For every cycle iteration new purchase transactions from the last seven days are added to the neural networks and new predictions are then generated.

#### 4. EXPERIMENTAL RESULTS

As mentioned in the previous section, data was collected in the form of purchase transactions and was then transformed to a time series of the number of days passed between two purchases. Initial experiments had one time series for each customer-product pair and the later experiments had one time series for each customer. As a consequence, at least two purchases were required for forming a time series and preferably more than that. The input to the network was the regularized time series with 3 time steps, which meant that only customer-product pairs/customers with at least 4 purchases could be used. Based on the minimal number of purchases per customer-product pair or per customer we created multiple subsets of the initial datasets: 6+ purchases, 11+ purchases, 16+ purchases, 20+ purchases, 50+ purchases and 100+ purchases dataset. For each of the experiments the test dataset consisted of purchases from the last six months of purchase transactions.

LSTM networks were implemented in Python using Keras library [20] running on top of TensorFlow [21]. Every network has one hidden layer with 50 cells with the relu activation function used. Due to the large number of networks there were to be trained and the long training time, training was done in 200 epochs for each network.

Result from experiments performed by training the LSTM networks with time series for customer-product pairs were obtained in a previous research and are shown in Table 1. Although the precision for predicted purchase increases when minimal number of customer-product pair increases, it is still relatively low meaning that there will be a lot of false positives. On the other hand, a great number of the realized purchases will be covered according to the higher recall. Additionally, both precision and recall are very high for the category “predicted no purchase”, i.e. in this case it is easier to determine those customer-product pairs where no purchase will be made.

**Table 1** Results of experiments done using univariate LSTM on customer-product level

	Predicted purchase		Predicted no purchase		MAE	MAPE
	Precision	Recall	Precision	Recall		
6+ purchases	12.76%	43.27%	96.82%	85.37%	90.31	140.03%
11+ purchases	15.89%	41.95%	94.93%	82.66%	67.98	107.49%
16+ purchases	20.22%	46.79%	93.21%	79.82%	58.27	96.75%
20+ purchases	23.49%	48.14%	91.37%	77.79%	52.16	88.71%
50+ purchases	43.97%	69.92%	77.40%	53.62%	44.28	81.33%
100+ purchases	62.20%	92.94%	70.00%	22.58%	35.05	57.56%

After performing these initial experiments and analyzing purchase transactions we decided to focus on predicting the next purchase on the customer level, taking in account all of their purchases regardless of the product. This resulted in a greater number of purchases in all time series and a lower number of neural networks to train. The input used for the neural networks was still the time series generated from the number of days

passed between two consecutive purchases. It can be noticed in the results (Table 2) that the precision for a predicted purchase category is significantly higher in all datasets, recall is also higher, and the same is true for the “predicted no purchase” category. Using these networks for predicting customers that will make a purchase could make it possible to devise actions oriented to such customers.

**Table 2** Results of experiments done using univariate LSTM on customer level

	Predicted purchase		Predicted no purchase		MAE	MAPE
	Precision	Recall	Precision	Recall		
6+ purchases	58.11%	78.62%	90.29%	77.82%	50.34	85.14%
11+ purchases	58.81%	78.98%	89.21%	75.85%	42.66	73.51%
16+ purchases	60.51%	79.75%	88.10%	74.23%	36.31	69.87%
20+ purchases	60.76%	79.21%	86.57%	72.38%	32.64	68.47%
50+ purchases	67.79%	84.18%	78.18%	58.64%	20.30	59.49%
100+ purchases	76.48%	95.10%	78.67%	38.19%	14.06	49.85%

The following two experiments were performed by adding features to input time series for LSTM neural networks. Besides using the number of days passed between two consecutive purchases, we first added the number of days from two purchases ago, and then also the number of days from three purchases ago. Although these time series have multiple features, both added features are derived from the original univariate time series and the resulting LSTM networks can be regarded as pseudo-multivariate LSTM neural networks. Experiments performed using pseudo-multivariate LSTM neural networks with one additional feature produced results shown in Table 3, while Table 4 contains results from pseudo-multivariate LSTM neural network with two additional features. Both setups produced increasingly better results than the previous ones for both categories. Precision is relatively constant even for datasets with a small minimal number of purchases per customer and this is perhaps the greatest achievement. Predicting customers that will make a purchase in the following seven days with a high certainty gives the vendor the possibility of performing a number of actions in terms of inventory planning and/or personalized advertising. While predictions based on customer-product purchases give more opportunity for inventory planning, due to the fact that a specific product is considered, aggregated customer purchases predictions are better suited for personalized advertising, i.e. offers with better buying conditions for multiple products user buys.

**Table 3** Results of experiments done using pseudo-multivariate LSTM on customer level with one additional feature

	Predicted purchase		Predicted no purchase		MAE	MAPE
	Precision	Recall	Precision	Recall		
6+ purchases	88.05%	86.27%	94.67%	95.42%	3.76	16.74%
11+ purchases	87.51%	85.86%	93.88%	94.65%	3.73	18.02%
16+ purchases	87.30%	87.30%	93.71%	93.71%	3.96	19.64%
20+ purchases	86.69%	87.52%	93.23%	92.74%	3.80	20.10%
50+ purchases	87.16%	88.79%	88.18%	86.48%	3.75	25.30%
100+ purchases	88.67%	95.87%	89.45%	74.11%	2.92	28.00%



**Table 4** Results of experiments done using pseudo-multivariate LSTM on customer level with two additional features

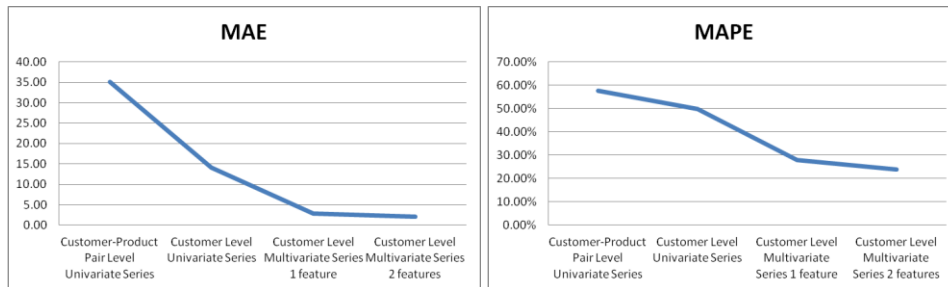
	Predicted purchase		Predicted no purchase		MAE	MAPE
	Precision	Recall	Precision	Recall		
6+ purchases	91.75%	88.20%	95.45%	96.90%	2.67	12.84%
11+ purchases	92.57%	89.26%	95.38%	96.87%	2.43	13.34%
16+ purchases	91.70%	88.63%	94.47%	96.03%	2.58	15.12%
20+ purchases	92.87%	88.20%	93.80%	96.34%	2.71	15.35%
50+ purchases	91.47%	90.77%	90.53%	91.25%	2.20	19.10%
100+ purchases	91.83%	94.64%	87.89%	82.20%	2.11	23.95%

Fig. 3 shows charts with precision and recall changes with each different network setup for four of the datasets. The first point in each chart represents values from initial experiments where each time series was formed for a specific customer-product pair. The second point represents results from univariate LSTM networks trained with customer level time series, while the third and fourth points represent results from pseudo-multivariate LSTM neural networks with one/two additional features. The prediction accuracy increase is apparent and drastic in datasets with a smaller minimal number of purchases per customer, but it is also noticeable in the dataset with 100+ purchases per customer.



**Fig. 3** Precision and recall of purchase prediction with different LSTM neural networks input for several datasets

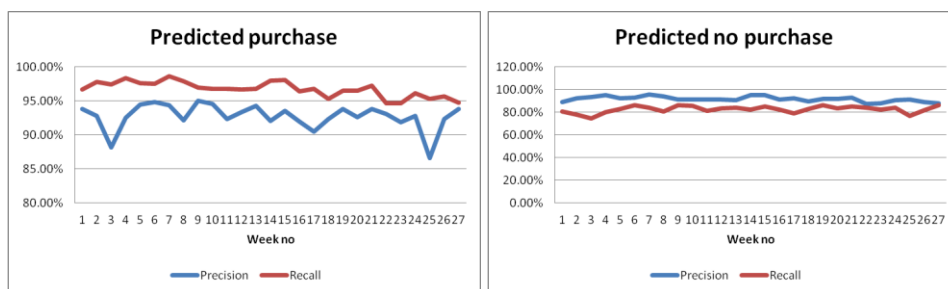
However, the improvement in accuracy for the dataset with the largest number of purchases per customer is perhaps best represented in Fig. 4. Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are metrics that are applied to the actual output values of the LSTM networks, before discretization to two categories, and both values show significant decrease with each improved neural network setup.



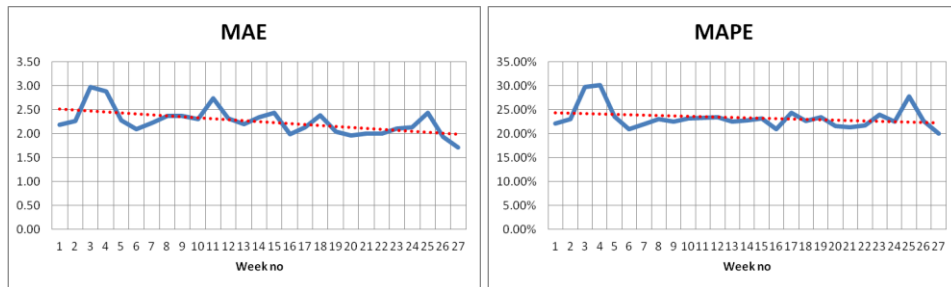
**Fig. 4** MAE and MAPE with different LSTM neural networks input for the dataset with at least 100 purchases per customer

After performing the first series of experiments, we adopted the final setup (pseudo-multivariate LSTM with two additional features) as the best performing one, based on the prediction results. In order to examine the consistency of results and the possible trend in metrics, we emulated a six month use of the LSTM for purchase prediction for the 100+ purchases dataset. Like in the previous experiments, we wanted to assure that the testing dataset always has at least six months of purchase transactions. This was achieved by starting at the date one year before the last purchase transactions. After training with the initial dataset and evaluating results, purchases from the following seven days were added to the LSTM networks and evaluation was repeated. The described process was reiterated 27 times in total, emulating a 27-week, i.e. six month use of the system.

Results of these experiments, i.e. changes of the evaluation metrics over time are shown in Fig. 5 and Fig. 6. For the output discretized to two categories it can be noticed that precision and recall are consistently high with some minor peaks. As for evaluation of the direct network output, MAE and MAPE vary in value a bit more, but for both a very slight downward trend is noticeable. This shows that over time, with additional input data, errors become smaller. With additional data following the original purchase transactions period, theoretically, MAE and MAPE could decrease even more.



**Fig. 5** Precision and recall during the 27-week testing period



**Fig. 6** MAE and MAPE value during the 27-week testing period

## 5. CONCLUSION

Long Short-Term Memory Neural Networks are a type of Recurrent Neural Networks that are especially convenient for predicting time series with long term dependencies which can be the case in purchasing transactions. Different researches have incorporated LSTM networks in purchase related predictions regarding multiple aspects of the purchase decisions. We focused on predicting when the next purchase will happen. Such information can be used by businesses for procurement planning, personalized advertising and other actions.

Collected data was transformed due to the irregular nature of the original time series and multiple series of experiments were performed. Results were evaluated in two forms: original values produced as network output and discretized to two classes. Predicting the next purchase for a customer-product pair gives very specific output information, but the accuracy of these predictions was relatively low. Switching to the customer level resulted in much higher precision and recall, and smaller errors. Finally, adding two additional features to the time series produced predictions with the highest accuracy and improved values of all metrics. It is important to note that the most significant improvement occurred in datasets with a small number of purchases per customer. The most successful network setup was tested emulating the intended use of the proposed system and the results were consistent during the entire period, with MAE and MAPE showing a slight downward trend.

Eliminating the product from the prediction target resulted in a great increase in accuracy, but has added a new variable to the problem. Depending on the type of products in the vendor company, it might not be necessary to predict exactly what product the customer wants to buy, but rather the product category. In a medical supply company as the data source, Unique Device Identification (UDI) can be used as helpful information in grouping products on one or more levels, simplifying the product prediction problem.

## REFERENCES

- [1] Y. Chai, G. Liu, Z. Chen, F. Li, Y. Li, and E. A. Effah, "A Temporal Collaborative Filtering Algorithm Based on Purchase Cycle," *Cloud Computing and Security*, vol. 11064, X. Sun, Z. Pan, and E. Bertino, Eds. Cham: Springer International Publishing, 2018, pp. 191–201, doi: 10.1007/978-3-030-00009-7\_18
- [2] J. Gruenen, C. Bode, and H. Hoehle, "Predictive Procurement Insights: B2B Business Network Contribution to Predictive Insights in the Procurement Process Following a Design Science Research Approach", *Designing the Digital Transformation*, vol 10243, Springer, 2017, pp. 267–281, doi: 10.1007/978-3-319-59144-5\_16.

- [3] M. Ćirić and B. Predić, “Predicting Purchase Day in B2B: From Statistical Methods towards LSTM Neural Networks,” Proceedings of the 10th International Conference on Information Society and Technology, Kopaonik, Serbia, 2020, ISBN: 978-86-85525-24-7, pp. 193-197.
- [4] B. Predić, N. Radosavljević, and A. Stojčić, “Time Series Analysis: Forecasting Sales Periods in Wholesale Systems” *Facta Univ. Ser. Autom. Control Robot.*, vol. 18, no. 3, Art. no. 3, Jan. 2020, doi: 10.22190/FUACR1903177P.
- [5] A. Verma, “Consumer Behaviour in Retail: Next Logical Purchase using Deep Neural Network” *ArXiv201006952 Cs*, Oct. 2020, Accessed: Nov. 11, 2020. [Online]. Available: <http://arxiv.org/abs/2010.06952>.
- [6] S.-M. Xie and C.-Y. Huang, “Systematic comparisons of customer base prediction accuracy: Pareto/NBD versus neural network” *Asia Pac. J. Mark. Logist.*, vol. ahead-of-print, no. ahead-of-print, May 2020, doi: 10.1108/APJML-09-2019-0520.
- [7] R. Gangurde, D. B. Kumar, and D. S. D. Gore, “Optimized Predictive Model using Artificial Neural Network for Market Basket Analysis”, *International Journal of Computer Science and Communication*, vol 9, issue 1, 2018, pp. 42-52, doi: 10.090592/IJCSC.2017.014
- [8] M. Hendriksen, E. Kuiper, P. Nauts, S. Schelter, and M. de Rijke, “Analyzing and Predicting Purchase Intent in E-commerce: Anonymous vs. Identified Customers”, SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, ISBN: 978-1-4503-8016-4, pp.
- [9] M. Korpusik, S. Sakaki, and F. C. Y.-Y. Chen, “Recurrent Neural Networks for Customer Purchase Prediction on Twitter”, *CBRecSys: Workshop on New Trends in Content-Based Recommender Systems at ACM Recommender Systems Conference*, 2016, ISSN: 1613-0073, pp. 47-50
- [10] D. Cirqueira, M. Helfert, and M. Bezbradica, “Towards Preprocessing Guidelines for Neural Network Embedding of Customer Behavior in Digital Retail”, *Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control*, Amsterdam Netherlands, Sep. 2019, pp. 1–6, doi: 10.1145/3386164.3389092.
- [11] A. Martínez, C. Schmuck, S. Pereverzyev, C. Pirker, and M. Haltmeier, “A machine learning framework for customer purchase prediction in the non-contractual setting” *Eur. J. Oper. Res.*, vol. 281, no. 3, pp. 588–596, Mar. 2020, doi: 10.1016/j.ejor.2018.04.034.
- [12] C. Huang *et al.*, “Online Purchase Prediction via Multi-Scale Modeling of Behavior Dynamics” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA, Jul. 2019, pp. 2613–2622, doi: 10.1145/3292500.3330790.
- [13] S. Stubseid and O. Arandjelovic, “Machine Learning Based Prediction of Consumer Purchasing Decisions: The Evidence and Its Significance”, *Proceeding of AAAI Conference on Artificial Intelligence*, Feb. 2018, ISBN 978-1-57735-801-5, pp. 100-106
- [14] G. Suchacka and S. Stemplewski, “Application of Neural Network to Predict Purchases in Online Store” in *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology – ISAT 2016 – Part IV*, vol. 524, Z. Wilimowska, L. Borzemski, A. Grzech, and J. Świątek, Eds. Cham: Springer International Publishing, 2017, ISBN: 978-3-319-46583-8, pp. 221–231.
- [15] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory” *Neural Comput.*, vol. 9, no. 8, , Nov. 1997, pp. 1735–1780, doi: 10.1162/neco.1997.9.8.1735.
- [16] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: continual prediction with LSTM”, 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), 1999, pp. 850–855, doi: 10.1049/cp:19991218.
- [17] S. Hochreiter, “The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions,” *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 06, no. 02, pp. 107–116, Apr. 1998, doi: 10.1142/S0218488598000094.
- [18] F. Gers, “Long Short-Term Memory in Recurrent Neural Networks,” PhD thesis, Ecole Polytechnique Federale De Lausanne, May 2001, doi: 10.5075/epfl-thesis-2366.
- [19] A. Graves, “Generating Sequences With Recurrent Neural Networks,” *ArXiv13080850 Cs*, Jun. 2014, Accessed: Nov. 11, 2020. [Online]. Available: <http://arxiv.org/abs/1308.0850>.
- [20] “Keras: the Python deep learning API.” <https://keras.io/> (accessed Nov. 15, 2020).
- [21] M. Abadi *et al.*, “TensorFlow: A system for large-scale machine learning”, Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, November 2016, ISBN: 978-1-931971-33-1, pp. 265–283