BÁNKI KÖZLEMÉNYEK
V. ÉVFOLYAM 1. SZÁM

Marketing using artificial neural networks

Anja Stijepović
University of Donja Gorica, Podgorica, Montenegro
anja.stijepovic2@udg.edu.me

Abstract — **Due to the emergence of numerous digital innovations, many business sectors are being reshaped at an astonishing pace. One of the technologies that is increasingly affecting every aspect of our lives is artificial intelligence. In this paper, we examine the impact of artificial intelligence in the context of marketing. The main application of artificial neural networks is in the field of predictive analytics, therefore these networks provide effective tools for predicting consumer behavior as well as for evaluating success of marketing campaigns. In order not to be overshadowed by their competition, companies and organizations must adopt new, fast and accurate methods of artificial intelligence when analysing their customer data. The artificial neural network model analysed in this paper is a prediction of the outflow of customers (churn) from a telecommunications company. Discovering the reasons for churn is very important in making future marketing decisions of a company. The accuracy of the resulting customer classification model could be increased by training on a balanced set of customer data.**

***Keywords*:** artificial neural networks, marketing, data sets, churn

1 INTRODUCTION

With the development of science and technology, our lives rely much more on digital products and services than they did in the previous decades. Numerous companies, aware of frequent technological innovations, are increasing automation and the use of artificial intelligence in everyday business. The technology of artificial intelligence is important because it enables human capabilities - understanding, reasoning, planning, communication and perception - to be performed by software more efficiently, effectively and at a low cost. Automation of these capabilities creates new opportunities in most business sectors and user applications.

Artificial neural networks are computer systems developed as a generalization of mathematical models of biological nervous systems in the human brain. Their ability to collect, memorize and use experimental knowledge (learn by example), influences their wide application. Artificial neural networks and machine learning are revolutionizing the world of marketing as well. Large amounts of data on customers, trends and products can now be processed faster and smarter in order to organize successful marketing campaigns. Neural networks, fed with enough data, are able to provide more accurate insights and predictions, helping marketers assess expectations and provide consumers with a better experience. Data analysis can answer the main challenge of digital advertising: „How to place an advertisement on the right platform and in front of relevant viewers?". Artificial neural networks provide companies with a deep insight into their customers and thus optimize the return on investment in advertising. Not only will companies be able to increasingly personalize interactions in the future, but they will also be able to predict future customer behavior based on the data collected.

2 ARTIFICIAL NEURAL NETWORKS

Following the example of biological nervous systems, which are composed of billions of neurons, artificial neural networks are created. Artificial neural networks consist of interconnected processing elements - artificial neurons [1]. With their help, computers are able to perform tasks that the human brain easily and successfully performs on a daily basis, such as recognizing shapes and voices, learning from examples, remembering and making decisions. Computers are machines that do routine work faster and better than humans. However, intelligence, awareness, creativity and feelings are human traits that computers lack. Therefore, artificial models inspired by the human brain allow computers to take over, facilitate or speed up many human tasks.

Artificial neurons have a simpler structure, but similar functions as biological ones. They, too, have a local memory in which they memorize the data they process and then pass it on through communication links. Using input data and weighting coefficients, artificial neurons calculate the output data, and then, with the help of some activation function, produce the final outputs. The 1. figure shows the architecture of an artificial neuron; $x_1, \ldots, x_4$ are the input data, $w_1, \ldots, w_4$ are weights, θ is the middle output, f is the activation function, and O is the final output [1]. The output signal of a neuron is given by the following relationship, where the variable net is defined as the scalar product of weight and input vectors:

$$O = f(net) = f\left(\sum_{j=1}^{n} \omega_j x_j\right). \quad (1)$$

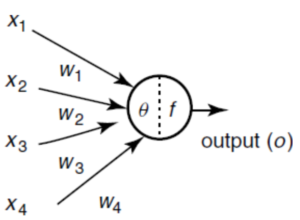

Figure 1: Architecture of an artificial neuron [1]

The activation function that is later used for producing inputs of the hidden layers is the ReLu function, defined by the following equation:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}. \quad (2)$$

The most common form of the activation function is the sigmoid function, which produces final output values between 0 and 1. The sigmoid function is defined by the following relationship:

$$f(net) = \frac{1}{1 + e^{-a \cdot net}}, \quad (3)$$

where the parameter a determines the slope of the function [2].

Adjusting the weights between neurons enables the learning ability of an artificial neuron [1]. Learning allows artificial neural networks to become more accurate in predicting outcomes. Networks are fed with a large amount of data which teaches them basic rules and improves their performance of various tasks. They are considered trained when a particular set of inputs leads to the desired final outputs.

The 2. figure shows an example of the artificial neural network architecture. We can conclude that artificial neural networks are graphs, whose nodes are neurons, and directed branches with weights represent the connections between inputs and outputs [4]. Usually, each layer of the network receives inputs from the previous layer or the environment, and sends its outputs to the next layer. Deep neural networks are composed by multiple hidden layers, and are used in deep learning (subset of machine learning).

3 How can artificial neural networks improve marketing?

In order to have a successful marketing campaign, a company must know its customers, that is, it must store data about them and then identify information relevant to further advertising of products and services. A company must observe the relationships with customers, find out where they are located, what age they are, which products they buy most often, the quantity and frequency of purchases, sensitivity to promotional activities and such [5]. All this data is stored in a large database of the company, which cannot be managed without the help of some computer software.

If we recall the ability of neural networks to learn from examples, it is clear that they can be taught to successfully analyse these types of data. The most common application of artificial neural networks is in the field of predictive analytics, which is an important part of creating a marketing plan for a business. Marketers can predict the future with a high probability and make the right decisions based on the expected results of advertising.

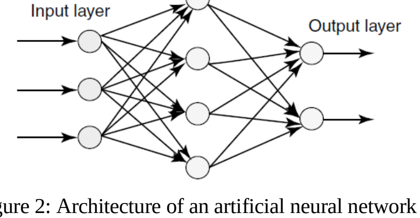

Figure 2: Architecture of an artificial neural network [1]

With the help of artificial intelligence, companies can optimize the return on investment in advertising by placing ads in front of relevant viewers. In addition to analysing simple data about custumers, like their name, address or age, patterns of behaviors should be considered during the formation of a marketing plan as well. In this context, hidden layer neurons can be viewed as invisible variables, which can be analysed using the weights of the connections between them and thus reveal important behaviors or attitudes [6].

Additionally, with the methods of artificial intelligence, it is possible to form propensity models. These models are designed to "identify potential customers who are more likely to respond to an offer" [7]. When creating a marketing campaign, the target audience and campaign goals are defined. Propensity models, by connecting customer characteristics with expected behavior, recommend strategies for achieving desired goals. They represent an opportunity to determine the value of a customer in the long run, i.e. how long the customer will remain loyal to the company. If a company has recognized its best customers, then it must make an effort to keep them [5].

So far, various applications of artificial neural networks in the field of marketing have been observed, including modeling consumer responses to market stimulations, new product development, sales forecasting, identifying consumer needs and expectations, market segmentation and creation of marketing strategies [6].

Another application of artificial neural networks is in the analysis of customer outflow (churn). One of the main rules that must be met during the formation of a company's marketing plan is that advertising to previous customers is cheaper and more effective than advertising to people who have not been in contact with the company before. [5] If a company loses valuable customers, it does not only lose the potential revenue from them, but it also must incur additional costs to replace old customers with new ones. The book [5] estimates that reducing customer outflows by only 5% per year could lead to a significant increase in profits (25%). Therefore, a company must pay special attention to retaining valuable customers whom the artificial intelligence model classified as „churned“ (likely to end transactions with a company). For example, telecommunications company T-Mobile analyses transactional data to determine valuable customers who are likely to change service providers. Then, T-mobile provides these customers with better offers and privileges in order to keep them in the company. [5] In the following research, we will analyse a model for predicting the outflow of customers of a telecommunications company.

4 Data collection and processing

The „Telco Customer Churn“ dataset found on the Kaggle website was used to create the classification model analysed in this research. Each row of this dataset represents a customer of one telecommunications company, and each column contains certain customer attributes. There are 21 columns and 704 customers (rows). The column named „Churn“ contains „Yes“ or „No“ answers depending on whether the customer has left the company in the previous month. It is important to determine whether a customer will leave the company, but also the reason for the potential churn, therefore we consider the „Churn“ column as a dependent variable (dependent on other columns). Other columns contain information about the services which each user signed up for (telephone, multiple lines, internet, online security, online backup, device protection, technical support and streaming of television and movies), account information (tenure, contract, method of payment, electronic payment, monthly and total charges), as well as demographic data on customers (gender, age and whether they have partners and dependents).

There are two types of data in this dataset: categorical and numerical. Most columns contain only Yes / No (or Female / Male) which are easily converted into numbers 1 or 0 and fed to the artificial neural network. The columns that posed a problem are „InternetService“, „Contract“ and „PaymentMethod“ which contain three or more textual responses. The problem was solved by the method of „one-hot encoding“, a process by which categorical variables are transformed into a form that can be given to a deep learning algorithm. In this way, each of these columns was divided into three or more columns, which now contain only 0 or 1. For example, in the initial column „Contract“, three unique answers were observed: „Month-to-month“, „One year” and „Two year”. A new dataset has been created which, instead of the „Contract“ column, contains „Month-to-month“, „One year“ and „Two year“ as separate columns with values of 0 (if the buyer does not have this type of contract) or 1 (otherwise). The numeric data from the „Tenure“, „MontlyCharges“, and „TotalCharges“ columns were scaled down to numbers between 0 and 1 so they can be comparable to other binary values in the dataset. Additionally, 11 rows which had no value for the total costs of the customer, but did for his monthly costs, were deleted from the initial dataset. The "customerID" column has been removed from the initial dataset because it does not affect the customer's decision whether to leave or not. In this way, data cleansing was performed before creating a deep learning model.

5 Methodology

Google Colab was used to analyse the data and create a predictive model, and the code was found on GitHub [3]. As Google Colab works in the cloud and supports many popular machine learning libraries, it has provided a suitable environment for model implementation. The implementation was done in Python, and the libraries used were Pandas, Numpy, Matplotlib, Scikit-learn and TenserFlow. Pandas and Numpy were used to sort and process the data, Matplotlib was used to visualize data via histograms, Scikit-learn was used to evaluate the performance of the machine learning algorithm and to calculate the interdependence of variables, and TenserFlow and Keras were used to create an artificial neural network.

After cleansing, the data was classified into two groups: the training group (80% of the data) and the testing group (the remaining 20%). The training dataset is used to fit into the model for its training, and the testing dataset is used to evaluate the machine learning model. The train-test split method was used to assess the performance of the created model over data that was not used in its training (data that the model had not seen before).

The artificial neural network was created with the Keras.Sequential command. The network had a fully-connected structure with an input layer, two hidden layers and an output layer. Fully connected layers were defined using the Dense class. The input shape was (26,) which represents the number of columns that the "Churn" column is depended on. That means the model expects rows of data with 26 variables. The first hidden layer had 20 neurons, the second hidden layer had 12, and the output layer had 1 neuron.The ReLu (rectified linear unit) activation function was used to obtain outputs from the hidden layers, and the sigmoid function was used for producing final output values between 0 and 1. The outputs were then transformed into numbers 0 or 1 for easier analysis. The output 0 means the model predicted the customer as not churned, and 1 as churned.

After the model was defined, it was compiled. Since churn prediction is a binary classification problem, cross entropy was used as the loss argument (defined in Keras as “binary_crossentropy“). The optimizer was defined as the efficient stochastic gradient descent algorithm “adam“. The classification accuracy was defined via the metrics argument. The model was then fitted and ran for 100 epochs.

Finally, with the classification report, the accuracy of the model with actual and predicted values was calculated. A summary of the results was presented by the confusion matrix, which showed us how many times the classifier was accurate on the testing dataset (how many customers were or were not classified as churned).

6 Discussion and results

During dataset analysis, it was noticed that most clients did not leave the company in the previous month, i.e. that the majority of values (73,42% of the total number of customers) in the „Churn“ column is equal to „No”. This fact is confirmed by the graph in the 3. figure, which shows the number of „Yes” and „No” answers in the „Churn” column. As the Yes/No response classes are not evenly distributed, the observed dataset is unbalanced, leading to the assumption that the model will not be successful in predicting customers who have left the company because it is trained with less data on such customers. The assumption was later confirmed by the classification report.

In order to discover the categorical values that have the biggest affect on the outflow of customers, mutual information was calculated (a measure of the interdependence between 15 independent variables and the dependent variable „Churn”). In the 4. figure we can see that the variable „Contract” has the highest mutual information (9,82%), therefore, among the observed variables, it most affects the outflow of customers. „Gender”, „PhoneService” and „MultipleLines” have mutual information that is close to 0, which means that these variables do not affect the outflow of customers. Variables that do not have a strong association with the target variable („Churn”) could be excluded from the creation of customer churn prediction models.

.

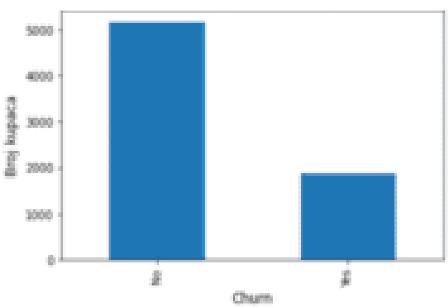

Figure 3: Graphic visualization of customer churn

Contract	0.098182
OnlineSecurity	0.064528
TechSupport	0.062873
InternetService	0.055394
OnlineBackup	0.046659
PaymentMethod	0.044423
DeviceProtection	0.043784
StreamingMovies	0.031918
StreamingTV	0.031803
PaperlessBilling	0.019119
Dependents	0.014270
Partner	0.011383
MultipleLines	0.000798
PhoneService	0.000069
gender	0.000037

Figure 4: Mutual information between categorical variables and the "Churn" variable

The „SeniorCitizen“ column was not considered in the previous analysis because it only contains numbers 0 and 1 in the initial dataset. Therefore, we separately analysed the variable „SeniorCitizen”, as well as the variable „Contract”, which showed the highest degree of dependence on the variable "Churn". The results are shown on the graphs in the 5. and 6. figure, where the red colour represents customers who are churned, and green represents those who are not.

In the 5. figure we can see that almost 80% of non-senior customers are not churned, which is less than 60% amongst senior customers. Thus, the outflow rate of senior citizens is almost twice as high as the outflow rate of young citizens. The company can use this information by giving older customers better offers that would prevent them from leaving the company. Based on the graph in the 6. figure, we conclude that clients with monthly contracts have higher outflow rates compared to clients with annual contracts.

Numerical variables „tenure” and „TotalCharges” were considered as well. In the following graphs shown in the 7. figure, the x −axis represents the number of months spent in the company (in the first graph) or the amount of total costs (in the second graph), while the y −axis represents the number of customers. We notice that the higher the total charges and the fewer months spent in the company, the greater the outflow of customers. This shows that loyal customers, i.e. customers who have been associated with the given telecommunications company for a long time, are unlikely to leave the company, whereas the opposite is true for newer customers. The company also needs to calculate whether it is more profitable to reduce the total costs of certain customers, or to lose those customers.

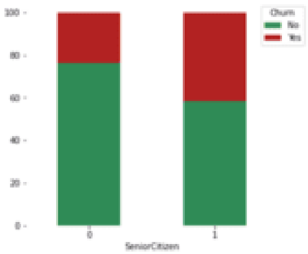

Figure 5: Churn rate of seniors and younger customers

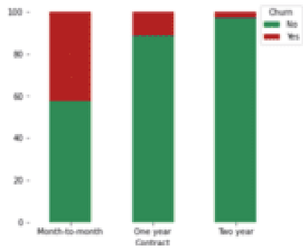

Figure 6: Churn rate of customers who have different types of contracts in the company

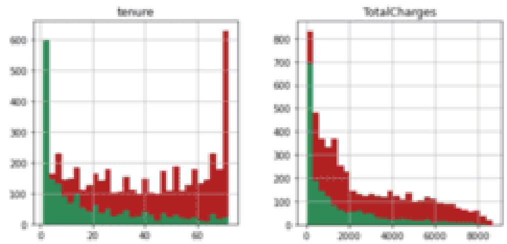

Figure 7: Customer churn depending on the number of months in the company and total charges (respectively)

After analysing the variables that could potentially affect the outflow, a model of artificial neural networks was created, which predicts the customer churn with a certain accuracy. The performance of the model was assessed by the classification report shown in the 8. figure.

The accuracy of the model (the ratio of correct predictions and all predictions) is 0.78. However, the accuracy of the model does not give enough information about its performance, because the model, as mentioned above, is built on an unbalanced data set. Therefore, we calculated separately the precision of the model in classifying customers who have not left the company (“churn” is 0), and the precision of the model in classifying customers who have left the company (“churn” is 1). Our assumption turned out to be correct, since the model is fed by a much larger number of data on customers who remained in the company, then the forecast for such customers is more accurate. This is confirmed by the classification report in which the accuracy for class 0 is equal to 0.83, while the accuracy for class 1 is equal to 0.64. Precision for class 1 is the ratio of the number of customers that the model correctly classified as churned, and the total number of customers it classified as churned. The recall for class 0 is the ratio of the number of customers that the model correctly classified as non-churned and the number of customers that are actually non-churned. As shown in the classification report, the recall for class 0 is 0.88, but the recall for class 1 is 0.55.

The „confusion matrix” shown in the 9. figure gives us information about the number of correct and incorrect classifications for 0 and 1. The columns of the matrix contain the predicted classes, while the rows represent the actual classes. So, out of 1407 data that entered the model’s testing group, 408 (185 + 223) were actually 1 (left the company), and 999 (875 + 124) were 0 (stayed in the company). However, the model predicted that 347 (223 + 124) of data is equal to 1, and 1060 (185 + 875) is equal to 0. Thus, the correct classifications of the model are on the main diagonal of the matrix, and incorrect are outside it.

	precision	recall	f1-score	support
0	0.83	0.88	0.85	999
1	0.64	0.55	0.59	408
accuracy			0.78	1407
macro avg	0.73	0.71	0.72	1407
weighted avg	0.77	0.78	0.77	1407

Figure 8: Classification report

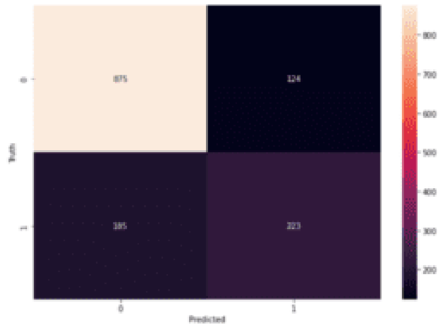

Figure 9: Confusion matrix

7 Conclusion

In this paper, the concept of artificial neural networks, as well as their application in marketing, were explained. The model of deep learning analysed in this paper solves the problem of classifying customers into groups of customers who have or have not left a particular company in the previous month. The model was trained with 80% of data from the given dataset and then tested on the remaining 20%. The model has proven successful in classifying customers who remain in the company, however, its accuracy in classifying customers who have left the company needs to be improved by further research. Better performance of the created model is expected by increasing the dataset (by balancing it), which would feed the neural network with enough of different data. Adding more hidden layers to the used neural network may improve the accuracy of the model. Of course, in order for a company to make the right marketing decisions, customer data must be further analysed statistically. Variables that do not have a strong correlation with the target variable „Churn” could be excluded from the creation of the classification model. Additionally, a successful model could be created by other methods of machine learning.

Outflow of customers is a problem that all companies face, however, it can be prevented by proper prediction. Analysis of big amounts of data can help companies detect valuable customers who plan to leave. After recognizing such customers, the company should strive to improve the services and products that those customers are dissatisfied with. Thus, the company must analyse the results of the successful predictive model in order to make successful marketing decisions and lose as few customers as possible.

References

[1] Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*, 901–903.

[2] Bašić, B. D., Čupić, M., Šnajder, J. (2008). Umjetne neuronske mreže. *Zagreb: Fakultet elektrotehnike i računarstva*, 7–15.

[3] Grover, K. (2021). *Deep-learning-keras-tf-tutorial/churn.ipynb at master · codebasics/deep-learning-keras-tf-tutorial.* Retrieved from https://github.com/codebasics/deep-learning-keras-tf-tutorial/blob/master/11_chrun_prediction/churn.ipynb. (Last accessed 13 May 2022).

[4] Jain, A. K., Mao, J., Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer* , *29* (3), 31–44.

[5] Lamb, C., Hair, J. F., McDaniel, C., Summers, J., Gardiner, M. (2009). *Mktg*. Cengage Learning Australia, 293–303.

[6] Silva, M., Moutinho, L. (2015). Artificial neural networks in marketing. *Wiley Encyclopedia of Management*, 1–5.

[7] Thiraviyam, T. (2018). *Artificial intelligence marketing*, 449–451.