Customer Churn Prediction in Telco Industry Using Artificial Neural Networks

Hristo Yanchev[1],

Assoc. Prof. Dr. Dorina Kabakchieva[2]

e-mail: hristoyanchev@yahoo.com, e-mail: dkabakchieva@unwe.bg

Abstract:

Customer churn is a well-known problem in many industries. The cost, in terms of money and time, for acquiring new customers is several times higher than retaining the existing ones. Therefore, developing a process in order to find these customers before they churn is crucial for the business, thus the company resources could be utilized for future projects instead of fulfilling clients shortage. Customer churn prediction is performed by carefully analyzing customer data including number of calls, length of calls, internet services used, tenure, monthly charges, technical support availability, etc. The effect of data normalization in an Artificial Neural Network model, applied to a dataset of 7043 customers in the telecom industry, is analyzed in this paper. Experiments with data normalization in an ANN model for finding potential customer churn, and the selection of training and testing partitions in the modelling phase, are conducted in the presented research. The achieved results reveal that data normalization is a must when using a Neural Network model and higher total accuracy doesn't mean higher class prediction percentage.

Keywords: Customer Churn, Churn Prediction, Telco Industry, Artificial Neural Network

JEL: L96, C38

1. Introduction

Customer churn is one of the major problems in many industries. Areas such as finance, healthcare, telecommunications, online gaming and others are affected. The problem is that finding new customers is many times more expensive than the keeping of current ones. However, companies realize this only after their customers have terminated their contracts and the company itself has analyzed the financial aspect of this problem. According to the CEO of Kovai, retaining 5% more customers can increase a company's revenue from 25% to 95% (Kumar, 2022). In this competitive world, companies are trying to use new innovative methods to find potential churners and focus their resources on them before these people leave. This report examines users in the telecommunications industry who tend to leave the company which they have a contract with.

In order to assess whether customers would switch a company, it is important to analyze their data. This is data that the company already has, such as number of calls, duration of calls, internet services used, technical support, etc.. Of course, there are many challenges before this data could become part of new churn detection model. The data can be incomplete - missing values, wrong values (e.g. in a field where numerical data is expected a nominal one to be present and vice versa), or the data can be in an inappropriate format. All these problems should be fixed and the data should be corrected to the expected data types (numeric, text, dates, etc.). Then, weights of the different fields might be calculated. Their values can be determined through

[1] PhD student, Department: Information Technologies and Communications, UNWE, e-mail: hristoyanchev@yahoo.com

[2] ASSOC. PROF, *Department: Information Technologies and Communications, UNWE*, e-mail: dkabakchieva@unwe.bg

detailed interviews with the people from the various departments of the company, whose data is part of the study, or by weighting the coefficients of these fields and checking for correlation. The next step is to train a model to flag potential churners. The last step is to decide which customers are important and it is worth to be retained.

The solution to this type of problems can be realized by using Data Mining methods and tools. This is a process in which various recurring patterns in the data can be detected, the presence of hidden relationships between the fields could be found, and thus the potential leaving customers of a given company could be detected. This type of problem is most often solved with a classification technique. In classification, the class of belonging of a given object is predicted. This object can belong to only one class and this is due to the combination of different parameters that the object possesses (Theodoridis, 2020).

2. State-of-the-art in the Field of Customer Churn Analysis

There are different types of customer churn – intentional and unintentional churn. Intentional churners are those who intentionally change their company and go to a competitor. Involuntary churners are those who terminate their contracts because they change their residence or didn't pay their monthly bills in time. The focus of this paper is on the intentional churners and their detection through Data Mining implemented with Artificial Neural Network model.

Neural networks are one of the most preferred models in Data Mining because of the high performance they offer. The main metric that will be tracked in the analyzed articles, when available, is the churn class prediction as the focus of this paper is on the churn detection.

In the article *Customer Churn Prediction Using Artificial Neural Network: An Analytical CRM Application* (Iranmanesh Seyed Hossein, 2019) the artificial neural network model is used to detect potential customers leaving the banking sector. The marked customers are further analyzed by being grouped into two groups - with high and low risk of dropping out and divided by their profession. After that, at-risk consumers are also grouped based on age groups: under 18, between 18 and 30, between 30 and 40, between 40 and 60, and over 60. Due to the two additional analyses, the authors present an opportunity for more precise targeting of customers, those whom the bank should pay attention to before they churn.

In the paper *Customer Churn Prediction Model Using ANN* (Baby Bestin, 2023), an ANN model is applied to the data that was analyzed and according to the results a precision rate of correctly tagged customers in the churn class of 79% and an overall model accuracy of 86% was achieved. Multiple metrics were then compared against the churn and loyal customers. Hence, article states that a relationship is established - people who have a larger range of bank products are more likely to remain loyal.

The paper titled *Customer Churn Prediction Using Ordinary Artificial Neural Network and Convolutional Neural Network Algorithms: A Comparative Performance Assessment* compares different churn detection models. The results from the analysis reveal that the artificial neural networks produce the best results, reaching 98.27% correct prediction of the dropout class (SEYMEN Omer Faruk, 2022). Additional fields describing the predicted group of users are not considered in this study.

According to the study presented in *Customer churn prediction model enhancement for the telecommunication industry using data transformation methods and feature selection* on users in the field of telecommunications (Boujelbene Zaineb, 2024), artificial neural networks reached 80.05% in predicting the churn class**.**

Artificial Neural Networks (ANN) or simply called Neural Networks (NN) are one of the most preferred models in Data Mining because of the high accuracy rate of the provided results. Neural networks are a mathematical model inspired by the human beings, and more specifically, the biological neural networks in the human brain. Artificial neural networks are composed of many simple nodes called neural cells or neurons (Baby Bestin, 2023). In most cases, ANNs can adapt and change their structure based on external or internal information that flows through the network during the training phase (SINGH YASHPAL, 2010).

Artificial neural networks consist of three main types of layers (Figure 1): input layer, hidden layers (one or more) and output layer. The input layer receives data that has already been prepared in an appropriate format. This is external information to the model and no calculations are performed in this layer. Neurons pass the information to the next layer. Cells in the Hidden layer receive information only from the input layer and have no access to the outside world. At this stage, numerous calculations are made based on the incoming data. Upon successful activation of the connections between the neurons and reaching a certain weight of the coefficients, information is sent to the Output layer. The Output layer displays the knowledge that the model has found (Geeksforgeeks, 2024).

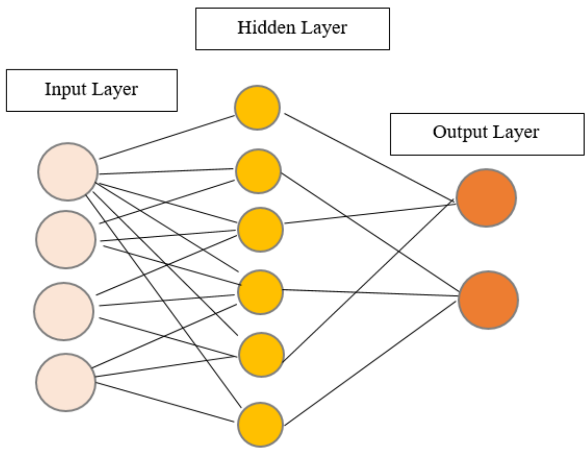

Figure 20. Neural Network

In the presented study, two of the components of the Data Mining process, using an artificial neural network model for detecting customer churn, are being changed, in order to improve the model and achieve better classification results.

3. Research Methodology

The conducted research is based on the CRISP-DM Model (Hotz, 2024) and includes the following steps (Figure 2):

- Business Understanding is the initial stage that focuses on understanding research objectives and formulating requirements from the perspective of business users. After the completion of the stage, the acquired knowledge is used to define the knowledge extraction tasks and to draw up a preliminary plan to achieve the objectives
- Data Understanding: This stage begins with collecting the necessary data and continues with activities aimed at deepening the researcher's knowledge of the nature of the data. At this stage, data quality issues need to be identified
- Data preparation includes all activities of creating from initial "raw" data the "final set" of data (data that will be used by the modeling tools). The stage of data preparation often has to be performed repeatedly, since the quality of the obtained results depends on the quality of the data.

- The modeling stage includes the selection and application of various modeling methods aimed at extracting knowledge from the data. The parameters of the models are calibrated to optimal values. Since some models have specific requirements regarding the data format, at this stage it is often necessary to return to the data preparation stage.
- Evaluation of the models is done with the aim of a deeper understanding of the created models from the point of view not only of the researcher, but also of the business users. It is important to carefully review all the steps involved in creating a particular model to ensure that they are achieving their intended goals.
- The already-made models can be used in two main ways. The analyst can recommend specific actions based on conclusions from the built model and the results obtained, or the model can be applied to new data. In the last decade, the CRISP-DM model has found extremely wide application both in business and among the scientific community, and has become a standard approach for the implementation of Data Mining projects.

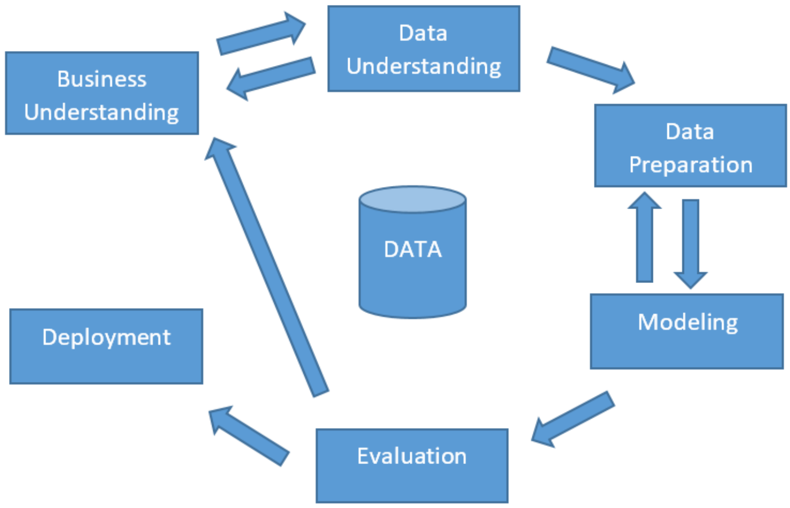

Figure 21. CRISP-DM Model

The software used for this research is RapidMiner.

Data preparation

In order to apply the selected model and to make tests for its improvement, it is necessary to first carefully analyze the data (table 1) on which it will be tested. The data is in excel format and includes 7043 rows (corresponding to 7043 customers) and 21 fields (the customer characteristic features).

Table 3. Attributes of the data set

Number	Attribute	Data type	Missing Values
1	customerID	Nominal	no
2	Gender	Nominal	no

3	SeniorCitizen	Nominal	no
4	Partner	Nominal	no
5	Dependents	Nominal	no
6	Tenure	Numeric	no
7	PhoneService	Nominal	no
8	MultipleLines	Nominal	no
9	InternetService	Nominal	no
10	OnlineSecurity	Nominal	no
11	OnlineBackup	Nominal	no
12	DeviceProtection	Nominal	no
13	TechSupport	Nominal	no
14	StreamingTV	Nominal	no
15	StreamingMovies	Nominal	no
16	Contract	Nominal	no
17	PaperlessBilling	Nominal	no
18	PaymentMethod	Nominal	no
19	MonthlyCharges	Numeric	no
20	TotalCharges	Numeric	yes
21	Churn	Nominal	no

All the data, with the exception of the customerID field, is relevant for the analysis. The customerID column contains information about the unique identifier of the customers, which is not related to their behavior and is not the reason for dropping or staying with the company.

Then the data is being transformed, missing values and incomplete fields removed. There is missing data in the TotalCharges field, in which case the corresponding rows where the value is missing is being removed in order not to negatively affect the model.

The data type conversion is also very important because most of the columns contain categorical data and the neural network model only works with numerical one. In the software used (RapidMiner), this conversion can be done in several ways. The “Dummy coding” option is selected, where all category data are converted to 1 or 0. For example The Online Security field contains three values - yes, no, no internet service. Using this type of conversion, three new fields are created: Online Security – yes, Online Security - no, Online Security – no internet service. Each row where the category value corresponds to the column name is filled with 1, and when the value is different than the column name, it is filled with 0.

Data normalization

Calculating the weights of the attributes to find the appropriate ones for the analysis is the next step. Information gain Ratio method is being used for this purpose. Through this method, the coefficients of the fields involved in the study are weighted, the higher the coefficient, the more significant the attribute to the model is. This method is recommended for categorical/nominal data type fields.

Comparison of results before and after applying data normalization is carried out.

Without normalization at this stage, the following results (table 2) are observed:

Table 4. Data without normalization

Attribute	Weight
gender	0,0001

PhoneService	0,0002
MultipleLines	0,0008
Partner	0,0164
Dependents	0,0234
SeniorCitizen	0,0237
PaperlessBilling	0,0283
StreamingTV	0,0299
StreamingMovies	0,0300
PaymentMethod	0,0325
DeviceProtection	0,0413
OnlineBackup	0,0440
MonthlyCharges	0,0494
InternetService	0,0523
TechSupport	0,0605
OnlineSecurity	0,0622
customerID	0,0654
tenure	0,0902
Contract	0,0984
TotalCharges	0,1344

The weights vary between 0.001 and 0.1344. Based on the weight indexes, attributes with a coefficient above 0.04 were selected without the participation of customerID, because this field does not carry significant value for detecting churned customers, it represents the unique identifier of a client.

Data normalization during the weighting of the coefficients leads to a change in the weight index of the attributes (table 3):

Table 5. Normalized data

Attribute	**Weight**
gender	0,0000
PhoneService	0,0012
MultipleLines	0,0059
Partner	0,1219
Dependents	0,1738
SeniorCitizen	0,1763
PaperlessBilling	0,2101
StreamingTV	0,2218
StreamingMovies	0,2226
PaymentMethod	0,2413
DeviceProtection	0,3072
OnlineBackup	0,3273

MonthlyCharges	0,3674
InternetService	0,3887
TechSupport	0,4501
OnlineSecurity	0,4628
customerID	0,4860
tenure	0,6710
Contract	0,7316
TotalCharges	1,0000

In this case, it is necessary to change the thresholds related to the attributes which will be included into the study. The coefficient index is changed from 0.04 to 0.3 in order not to include all attributes into the model. Therefore, all attributes with a coefficient above 0.3 were selected for the modelling phase.

4. *Achieved Results*

Experimenting with Data Normalization

The results from the evaluation of the model accuracy with pre-normalization of the data when weighting the attribute coefficients are presented in Table 4

Table 6. Model Accuracy Evaluation after pre-normalization

	True No	True Yes	Class prediction
Pred. No	1878	377	83.28%
Pred. Yes	187	371	66.49%
Class Recall	90.94%	49.60%	

The accuracy of the model with no applied data normalization before weighting the attribute coefficients did not change at all compared to the previous results (table 5).

Table 7. Model Accuracy Evaluation after normalization during the weights are calculated

	True No	True Yes	Class prediction
Pred. No	1878	377	83.28%
Pred. Yes	187	371	66.49%
Class Recall	90.94%	49.60%	

Data normalization can also be done in the neural network model itself. The following results (table 6) were achieved when the data was not normalized. Both of the results presented above had data normalization enabled into the ANN model.

Table 8. Results after applying ANN without any normalization of data

	True No	True Yes	Class prediction
Pred. No	2065	748	73.41%
Pred. Yes	0	0	0.00%
Class Recall	100.00%	0.00%	

The results show that the neural network model cannot work correctly without data normalization. The accuracy of the model drops to 73% from 80%, and the correct prediction of the class responsible for leaving customers (Class prediction Churn) is 0%, which excludes the possibility of solving the business problem in this way.

Experimenting with the ratio of the training and testing data partitions

The next criterion that was investigated in order to improve the model was the ratio split of the training and testing data. The conducted research on the selection of training and testing data, summarized in Table 7, refers to the percentage of data used for training and testing of the applied model.

Table 9. Comparison of the split for training and testing data in the modeling phase of the data mining process

Source	training data split	test data split
(SEYMEN Omer Faruk, 2022)	70 %	30 %
(Baby Bestin, 2023), (Thangeda Rahul, 2024), (Rudd D. H., 2021), (Boujelbene Zaineb, 2024) и (Mouli Kathi Chandra, 2024)	80 %	20 %
(Geiler Louis, 2022)	75 %	25 %

The results shown in tables 4 and 5 were obtained by splitting the data into 60% for training and 40% for testing the model. When 70% of the data is used for training, the overall accuracy of 79.53% is reached for the ANN model, when 60% of the data was used for training the overall accuracy was 79.95%. There is a visible decrease in the overall accuracy of the model and a decrease in the percentage of the churn class prediction - 64.30% (table 8) is achieved, compared to the results displayed in table 4 and table 5 where 60% of the data was used to train the model.

Table 10. Results obtained for 70% of the data used for training of the model

	True No	True Yes	Class prediction
Pred. No	1388	271	83.98%
Pred. Yes	161	290	64.30%
Class Recall	89.61%	51.69%	

After a change in the amount of data allocated for training the model to 80% was made, the total accuracy increased to 80.03% (table 9) and there was an increase in the percentage of prediction class.

Table 11. Results after 80% of the data is split for training the model

	True No	True Yes	Class prediction
Pred. No	928	176	84.06%
Pred. Yes	105	198	65.35%
Class Recall	89.84%	52.94%	

Finally after the volume of data allocated for training the model to 75% was made the total accuracy reached 79.24%, which is a drop of 0.8 percent in the overall accuracy of the model and there was a decrease in the percentage predicting the churn class reaching 64.01% (table 10).

Table 12. Results after 75% of the data is split for training the model

	True No	True Yes	Class prediction
Pred. No	1160	234	83.21%
Pred. Yes	131	233	64.01%
Class Recall	89.85%	49.89%	

The final results in this study are summarized into table 11. The data are sorted by the criteria of data normalization, training partition volume and class prediction – churn percentage.

Table 13. Results summary table

Criteria	Training partition volume	Class prediction - Yes
Pre-weighting normalization plus data normalization into the model	60%	66.29%
Normalization during weighting plus data normalization into the model	60%	66.29%
Without any data normalization	60%	00.00%
Data normalized into the ANN model	70%	64.30%
Data normalized into the ANN model	75%	64.01%
Data normalized into the ANN model	80%	65.35%

Conclusion

One of the most common problems that all service providing companies face is losing customers. Identifying these customers in a timely manner is a challenging task with no universal solution. Therefore, it's essential to thoroughly analyze each case and implement a tailored approach based on the findings.

ANNs offer a high accuracy of predicting the target group of customers, but achieving better results is related to experimentation with the various components of the developed process. In this paper, the best results achieved were not related to allocating the largest sample for model training. The highest Churn class prediction rate was 66.29% and it was obtained by splitting the data 60% for training and 40% for testing in the developed model.

In the next phase of the research, to further improve the identification of churned customers and enhance model accuracy, it will be essential to group and analyze flagged users based on demographic information, service data, and consumption patterns. This approach will lead to easier identification of these customers, focusing on the right groups to prevent churn before it occurs.

References

Baby Bestin, Z. D. (2023). Customer Churn Prediction Model Using ANN.

Boujelbene Zaineb, M. Y. (2024). Customer churn prediction model enhacement for the telecommunication industry using data transformation methods and feature selection. www.globalscientificjournal.com.

Geeksforgeeks. (2024, June). Activation functions in Neural Networks. Retrieved from geeksforgeeks: https://www.geeksforgeeks.org/activation-functions-neural-networks/

Geiler Louis, S. A. (2022). A survey on machine learning methods for churn prediction. International Journal of Data Science and Analytics, 217-242.

Iranmanesh Seyed Hossein, M. H. (2019). Customer Churn Prediction Using Artificial Neural Network: An Analytical CRM Application. Industrial Engineering and Operations Management. Pilsen: Czech Republic.

Kumar, S. (2022). Retrieved from https://www.forbes.com/councils/forbesbusinesscouncil/2022/12/12/customer-retention-versus-customer-acquisition/

Mouli Kathi Chandra, C. V. (2024). An analysis on classification models for customer churn prediction. Cogent Engineering.

Rudd D. H., H. H. (2021). Causal Analysis of Customer Churn Using Deep Learning. 2021 International Conference on Digital Society and Intelligent Systems (DSInS), (pp. 319-324). Chengdu.

SEYMEN Omer Faruk, E. O. (2022). Customer Churn Prediction Using Ordinary Artificial Neural Network and Convolutional Neural Network Algorithms: A Comparative Performance Assessment. Journal of Science, 720-733.

SINGH YASHPAL, A. S. (2010). NEURAL NETWORKS IN DATA MINING.

Thangeda Rahul, N. K. (2024). A neural network-based predictive decision model for customer retention in the telecommunication sector. Technological Forecasting & Social Change.

Theodoridis, S. (2020). Chapter 3 - Learning in Parametric Modeling: Basic Concepts and Directions. In S. Theodoridis, Machine Learning (pp. 67-120). Retrieved from https://www.sciencedirect.com/science/article/abs/pii/B978012818803300012X