A comparison between DSS and ML models for churn prediction

Boris Delibašić [1*], Sandro Radovanović [1], Marko Bohanec [2], and Milija Suknović[1]

[1] University of Belgrade, Faculty of Organizational Sciences, Belgrade, Serbia
[2] Jožef Stefan Institute, Department of Knowledge Technologies, Ljubljana, Slovenia
[*] boris.delibasic@fon.bg.ac.rs

Abstract: This paper compares the accuracy and convenience of a classical machine learning algorithm, a decision tree, and a classical decision support system model, built by the DEX (Decision EXpert) multicriteria decision modelling method for categorical data, on a churn prediction data set. Decision support systems (DSS) are a technology from the 1960s that was predominantly overruled by machine learning (ML) in the 2010s due to the explosion of big data, and their cost effectiveness. Here we discuss the similar and different aspects of the two technologies, and demonstrate the performance of these different, yet intertwined technologies. We show that our proposed DSS model outperforms the ML model.

Keywords: Churn Prediction; DSS; Multi-Criteria Models, DEX; DIDEX; Decision Tree; Machine Learning

1. Introduction

Machine Learning (ML) and Decision Support Systems (DSS) are technologies that are built for the same purpose, to support decision-making. While ML systems are fuelled by data, it is the expert's knowledge that builds DSS. The problem this paper is addressing is which of these systems is more suitable for customer churn prediction.

Customer churn is a global problem occurring in many domains, from university education students, clients in telecommunication companies, to employees in companies worldwide. Both DSS and ML have been used for customer churn prediction with reported advantage of ML use over the DSS technology.

Although it is not easy to compare different approaches, as usually a multidimensional perspective on a problem boils down to comparing a single criterion (like accuracy, or generalization error). Here, we adopt a DSS perspective for comparing the models using intrinsic criteria proposed by (Bohanec, 2021). We also chose to compare an ML model (decision tree), a DSS model developed by experts (DEX), and a DSS model generated by ML (DIDEX).

The DSS criteria that will support our discussion, as proposed by (Bohanec, 2021), are:

1. Correctness: a model offered should be accurate but also aligned with the problem that should be solved.
2. Completeness: a model should be able to work on all possible input value ranges.
3. Consistency: a model should be internally consistent and should provide consistent solutions.
4. Comprehensibility: a model should be understandable and transparent to the decision makers.
5. Convenience: a model should be easy to use and possess convenient properties, like sensitivity analysis and counterfactual reasoning.

We will show that the five criteria are to some extent present in the ML model, where the DSS model possesses them *per se*. The remainder of this short paper is structured as follows: In Section 2 we show similar background research, in Section 3 we demonstrate the experiment, and we make our conclusions in Section 4.

2. Background research

Customer churn prediction is a critical challenge in various industries, particularly telecommunications. ML techniques have been widely applied to address this issue. Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and Neural Networks are commonly used for churn prediction (De et al., 2021; Raj et al., 2024). These methods have shown promising results in terms of accuracy, precision, recall, and F1 score (Raj et al., 2024). While hybrid and ensemble methods have improved model performance, authors propose future research directions to include exploring deep learning techniques for enhanced prediction accuracy and personalization (Raj et al., 2024). Britto & Gobinath (2020) claim that predictive analytics in customer churn prediction offer more accurate outcomes compared to other approaches.

Within the domain of DSS, several studies have explored the use of rough set theory (RST) for customer churn prediction in the telecommunications industry. Amin et al. (2014, 2015, 2017) consistently found that RST-based approaches, particularly those using genetic algorithms for rule generation, outperformed other rule generation methods such as exhaustive, covering, and LEM2 algorithms. These studies demonstrated that RST can effectively classify and predict customer churn, providing valuable insights for strategic decision-making (Amin et al., 2015, 2017). The researchers also compared RST as a one-class and multi-class classifier, with the latter showing superior performance in binary and multi-class classification problems (Amin et al., 2014). Furthermore, attribute-level analysis using RST was found to be beneficial in developing customer retention policies (Amin et al., 2017). Overall, these studies highlight the potential of RST as an efficient, rule-based approach for customer churn prediction in the telecommunications sector, offering a globally optimal solution when compared to other state-of-the-art methods.

Fuzzy set theory has been applied to customer churn prediction in various industries. In retail banking, fuzzy c-means clustering was used to develop churn prediction models, outperforming classical methods (Popovic & Basic, 2008; Popovic & Basic, 2009). In the finance sector, a three-phase framework incorporating fuzzy inference systems was proposed to predict churn among high-value customers (Safinejad et al., 2018). For telecommunications companies, a model combining fuzzy logic and neural networks was developed to improve churn prediction accuracy (Papa et al., 2021). This method utilized normalized data to form membership function parameters and employed neural network algorithms, demonstrating the potential of fuzzy neural networks in customer churn forecasting. A DIDEX model has been applied to public policy decision-making in the Republic of Serbia (Delibašić et al, 2023). While rule-based methods excel in interpretability, some struggle to provide prediction probabilities, which are valuable for customer evaluation (Huang et al., 2011). Overall, rule-based systems offer a balance between accuracy and explainability in churn prediction.

Based on the previous work, we can conclude that while ML algorithms excel in accuracy, rule-based algorithms excel in interpretability. In this short paper we will demonstrate that DSS models can be both accurate and interpretable.

3. The experiment

For the experiment in this short paper a churn dataset[18] was used. The dataset contained 3333 rows, 20 attributes from which one attribute was the binary outcome churn attribute. The experiment was conducted in Orange software[19], and DEXi software[20]. The goal was to build a prediction model that can "*Identify intention of clients to churn in future, so that there is a possibility to react and possibly prevent churn*".

The experiment had the following consecution:

1. An ML model, namely a decision tree, was built on the churn dataset in Orange.
2. Two DSS models were developed in DEXi for the same purpose, one "handcrafted" by an expert

[18] https://github.com/albayraktaroglu/Datasets/blob/master/churn.csv
[19] https://orangedatamining.com/
[20] https://kt.ijs.si/MarkoBohanec/dexi.html

and the other build on the churn dataset using the method DIDEX

3. All DSS model was evaluated on the churn dataset.
4. A what-if analysis on the churn dataset was performed with the DSS model.
5. The “best” model was chosen according to the criteria proposed above and the results discussed.

3.1 The ML model

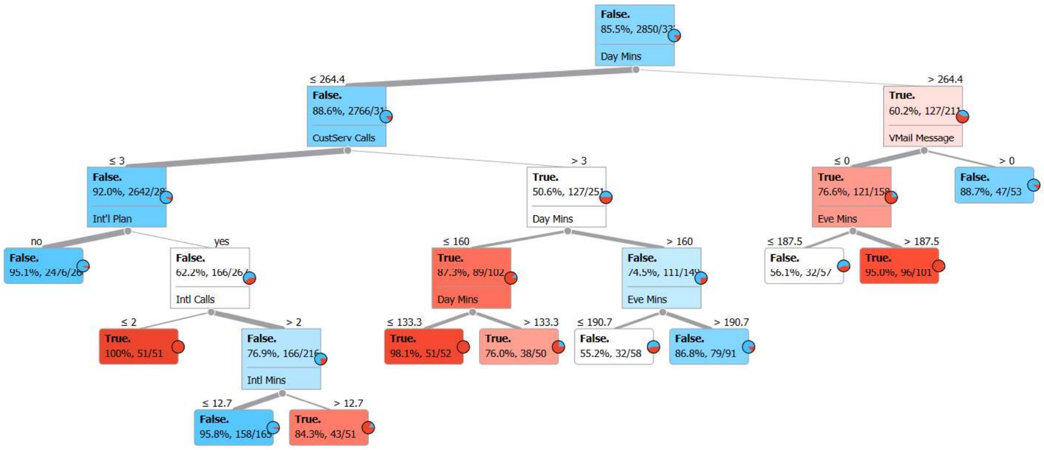


Figure 2: The churn prediction decision tree machine learned model

The decision tree, shown in Figure 1, has been grown in Orange software with limiting the depth of the decision tree to 6, as the generalization error was lowest with this depth. The decision tree model achieved the following performance: Accuracy 90.7%, Recall 56.7%, Precision 73.3%, and F1 63.9%. The tree was built using 10-fold cross validation. The following attributes were chosen as the most important from the 18 available in the dataset (without considering the *Phone number* attribute as the unique client id):

1. *Day Minutes* (how many minutes during a day a client made in a period)
2. *Customer Service Calls* (how many times a client called the customer service)
3. *Voice Mail Messages* (how many times a client left a voice mail message)
4. *Eve Minutes* (how many minutes during an evening a client made in a period)
5. *International Plan* (whether a client used the international plan package)
6. *International Calls* (how many international calls a client made in a period)
7. *International Minutes* (how many international minutes a client made in a period)

3.2 The First DSS model

Two DSS models were developed. The first was built in DEXi software, i.e. it was handcrafted by an expert in DEX methodology (shown in Figure 2). Six attributes were used for building the model. These were the 6 attributes identified in 3.1, excluding attribute 7. Attributes were discretized using the bounds in Figure 1. The first DEX model has a flat attribute structure, meaning that all attributes were on the same level of hierarchy w.r.t. to the root attribute [Churn].

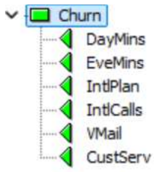
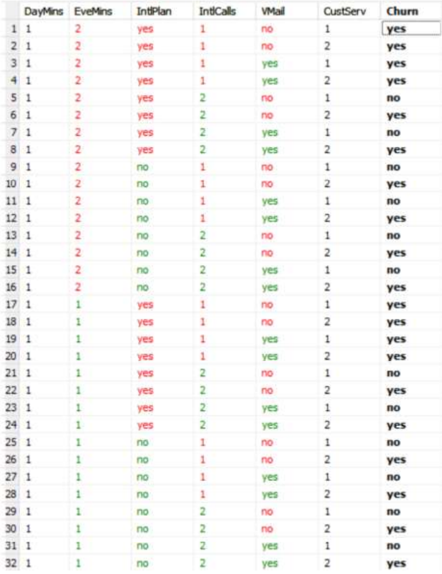


Figure 3: The expert modeled DSS using DEX methodology

96 rules were identified combining all categories in the six attributes and making a prediction for the churn outcome. A part of the decision rules is shown in Figure 3. Row 1 is read like this: If *Day Minutes* are Very Low (1), and *Eve Minutes* High (2), and *International Plan* is “yes”, *International Calls* is Low, *Voice Mail* is “No”, and *Customer Service Calls* is “Low” then [*Churn*] is “yes”.

The 96 rules have been applied to the whole dataset, i.e. to the 3333 rows and the following performance of the DSS model were achieved: Accuracy 74.71%, Recall 62.94%, Precision 31.4%, and F1 41.9%.



	DayMins	EveMins	IntlPlan	IntlCalls	VMail	CustServ	Churn
1	1	2	yes	1	no	1	yes
2	1	2	yes	1	no	2	yes
3	1	2	yes	1	yes	1	yes
4	1	2	yes	1	yes	2	yes
5	1	2	yes	2	no	1	no
6	1	2	yes	2	no	2	yes
7	1	2	yes	2	yes	1	no
8	1	2	yes	2	yes	2	yes
9	1	2	no	1	no	1	no
10	1	2	no	1	no	2	yes
11	1	2	no	1	yes	1	no
12	1	2	no	1	yes	2	yes
13	1	2	no	2	no	1	no
14	1	2	no	2	no	2	yes
15	1	2	no	2	yes	1	no
16	1	2	no	2	yes	2	yes
17	1	1	yes	1	no	1	yes
18	1	1	yes	1	no	2	yes
19	1	1	yes	1	yes	1	yes
20	1	1	yes	1	yes	2	yes
21	1	1	yes	2	no	1	no
22	1	1	yes	2	no	2	yes
23	1	1	yes	2	yes	1	no
24	1	1	yes	2	yes	2	yes
25	1	1	no	1	no	1	no
26	1	1	no	1	no	2	yes
27	1	1	no	1	yes	1	no
28	1	1	no	1	yes	2	yes
29	1	1	no	2	no	1	no
30	1	1	no	2	no	2	yes
31	1	1	no	2	yes	1	no
32	1	1	no	2	yes	2	yes

Figure 4: A snippet of expert-modeled decision rules using the DEX method

A what-if analysis was performed (called ± Analysis in DEXi), a task that is uncommon usingML models, but intrinsically supported in DSS models.

In Figure 4 two clients (rows 6, and 3328) were chosen for the analysis. They are both identified as churners. The analysis reveals insights what intervention could be done with this client in order that the client changes his/her mind. In the figure an adverse decision (no) can be seen five times, three times with client 6, and two times with client 3328. Client 6 would have a different churn outcome if: The number of *Day Minutes* would drop from category 3 (high) to category 2 (low), *Eve Minutes* would drop from category 2 (high) to category 1 (low), and if *Voice Mail* would change from category no to category yes. Client 3328 would change the outcome in two cases: if *Day Minutes* would increase from category 1 (very low) to 2 (low), and if *Customer Service calls* would change from category 2 (high) to category 1 (low).

Attribute	-1	6	+1
Churn		yes	
DayMins	no	3	]
EveMins	[	2	no
IntlPlan	[	yes	
IntlCalls		2	]
VMail	[	no	no
CustServ	[	1	

Attribute	-1	3328	+1
Churn		yes	
DayMins	[	1	no
EveMins		1	]
IntlPlan		no	]
IntlCalls		2	]
VMail	[	no	
CustServ	no	2	]

Figure 5: ±1 Analysis for clients 6 and 3328

3.3 The Second DSS model

The second DSS model was extracted using the DIDEX method (Radovanović et al. 2023), which automatically builds a DSS model from data. The difference with the DEX method is that in DEX method data does not exist, and experts create the attribute hierarchy and the decision rules by themselves. DIDEX extracts the attribute hierarchy and decision rules automatically from data. Only the names of newly constructed attributes are proposed by the user. The second DSS model is shown in Figure 5. After being learned from data, the model was read into DEXi software.

The attributes *International Plan* and *Voice Mail Plan* are grouped under a newly established feature [*Plans*], *International Calls* and *Customer Service Calls* are grouped under a new feature [*Calls*]. *Day Minutes* and *Evening Minutes* are grouped under a new feature [*Minutes*], where the features [*Calls*] and [*Minutes*] are grouped in a new feature [*Minutes & Calls*]. Finally, the decision on [*Churn*] is made based on features [*Plans*] and [*Minutes & Calls*].

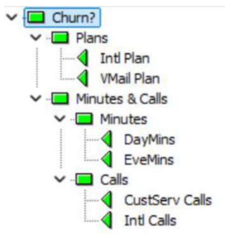


Figure 6: The expert modeled DSS using DEX method

A total of 22 decision rules were extracted using DIDEX method. From left to right, as shown in Figure 6, these rules are used to generate decision outcomes for the features [*Churn*], [*Plans*], [*Minutes & Calls*], [*Minutes*], and [*Calls*].

	Plans	Minutes & C	Churn?
1	0	0	No
2	0	1	Yes
3	1	0	Yes
4	1	1	Yes

	Intl Plan	VMail Plan	Plans
1	no	no	0
2	no	yes	0
3	yes	no	1
4	yes	yes	1

	Minutes	Calls	Minutes & Calls
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	1

	DayMins	EveMins	Minutes
1	1	1	0
2	1	2	0
3	2	1	0
4	2	2	0
5	3	1	0
5	3	2	1

	CustServ Calls	Intl Calls	Calls
1	1	1	0
2	1	2	1
3	2	1	0
4	2	2	0

Figure 7: The extracted rules using DIDEX method

The 22 extracted rules were again applied to the whole dataset, and the following performance of the DSS model were achieved: Accuracy 64.66%, Recall 79.09%, Precision 26.18%, and F1 39.34%.

The what-if analysis was then performed on the client 3333 (Figure 7). This time a comparison was made between the first DSS (DEX expert modelled) model, and the second DSS (DIDEX extracted) model.

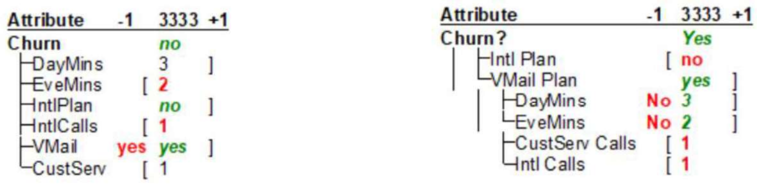


Figure 8: ±1 Sensitivity Analysis for clients 6 and 3328

It can be noticed that the two DSS models produce different churn predictions and have different recommendations on what to do if the churn decision should be different. While the first model thinks that the client 3333 will not churn and would do this in case where he/she would stop using voice mail, the second model thinks that the costumer would leave, and this outcome would have changed if either *Day Minutes* or *Eve Minutes* were reduced.

3.4 Which model to chose

It can be noticed that the ML model had the highest accuracy, precision and F1 value, and the poorest recall. The DSS DEX model achieved lower values from the ML model for all performance measures but recall, which was higher. The DSS DIDEX model had the lowest accuracy, precision, and F1 values, but the highest recall. The question is which model would be the most adequate.

According to Taskin (2023) in the case of telecom churn a false negative cost between 5 and 7 times more than a false positive, so the ranking of the models is:

1. DSS DEX model
2. DSS DIDEX model
3. ML model

This solution is stable on the whole interval [5,7]. The confusion matrices for the three models are shown in Table 1.

Table 6. Confusion matrices for DEX DSS model, DIDEX DSS model, and ML model

DEX	No	Yes	**DIDEX**	No	Yes	**ML**	No	Yes
No	2186	664	No	1765	1085	No	2750	100
Yes	179	304	Yes	173	310	Yes	209	274

While this result was highly unexpected, it was shown on this example of customer churn that a fully crafted DSS model can outperform, in some cases, ML models. In this example the recall of the DSS DEX model was the most adequate for tackling the costs of false classification. Having in mind the other DSS properties being satisfied, and the what-if analysis possibilities of the DSS models, the DSS model produced by DEX would be the most adequate choice in this situation.

4. Conclusions

In the 2010s, ML outperformed and mainly replaced DSSs due to a large availability of big data, availability of bigger processing power, easier integration into business processes, among others. ML does not require deep modeling from domain experts and easily overcomes their objective limitations.

Off-the-shelf ML algorithms most often optimize for accuracy, which may not be aligned with business goals. Even though users can influence this through decision thresholds or cost matrices, one needs to test whether these adjustments achieve the desired outcomes. Therefore, the property Correctness can easily be dissatisfied. In addition, if the available data is not representative of the entire population but is instead a biased sample (due to representation or sampling bias), the resulting model will inherit and potentially amplify those biases, leading to unfair or unreliable predictions which endangers Consistency. This property is recently handled according to “responsible AI systems” that are fair, and transparent. The first-generation ML algorithms would easily fail the consistency test of DSSs. Comprehensibility is a feature that is recently tackled by explainable AI systems, where previously ML systems were often regarded as black-boxes, and this was acceptable due to their high accuracy performances. The property Convenience of the DSS is just partially satisfied by ML systems, as ML systems do have a lot of convenient properties, which made them the dominant AI technology. Easy integration into business processes, and automation of those, by lowering the costs of business processes significantly, made ML models the dominant technology. Still, intrinsic properties of DSS systems, like sensitivity analysis, are not easily carried out in ML systems.

We demonstrated in this short paper the strength of DSS systems, which, although lost their previous glory, still can produce valuable models for the decision-making process and in-depth decision analysis and can be a very respectable alternative to ML and AI systems.

References

Amin, A., Khan, C., Ali, I., & Anwar, S. (2014). Customer churn prediction in telecommunication industry: With and without counter-example. In Nature-Inspired Computation and Machine Learning: 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16-22, 2014. Proceedings, Part II 13 (pp. 206-218). Springer International Publishing.

Amin, A., Shehzad, S., Khan, C., Ali, I., & Anwar, S. (2015). Churn prediction in telecommunication industry using rough set approach. New trends in computational collective intelligence, 83-95.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing, 237, 242-254.

Bohanec, M. (2021). From data and models to decision support systems: Lessons and advice for the future. In *EURO Working Group on DSS: A tour of the DSS developments over the last 30 years* (pp. 191-211). Cham: Springer International Publishing.

Britto, J. Gobinath R (2020).“A Detailed Review For Marketing Decision Making Support System In A Customer Churn Prediction,”. *Int. J. Sci. Technol. Res*, *9*(4), 3698-3702.

De, S., Prabu, P., & Paulose, J. (2021, September). Effective ML techniques to predict customer churn. In 2021 Third international conference on inventive research in computing applications (ICIRCA) (pp. 895-902). IEEE.

Delibašić, B., Radovanović, S., & Vukanović, S. (2023). A Decision Support System for Internal Migration Policy-Making. https://ipsitransactions.org/journals/papers/tir/2023jul/p7.pdf

Huang, Y., Huang, B., & Kechadi, M. T. (2011, May). A rule-based method for customer churn prediction in telecommunication services. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 411-422). Berlin, Heidelberg: Springer Berlin Heidelberg.

Papa, A., Shemet, Y., Yarovyi, A. (2021). Analysis of fuzzy logic methods for forecasting customer churn. Technology Audit and Production Reserves, 1 (2 (57)), 12–14. doi: http://doi.org/ 10.15587/2706-5448.2021.225285

Popović, D., & Bašić, B. D. (2008). Churn prediction model in retail banking using fuzzy c-means clustering. In Proceedings of the Conference on Data Mining and Data Warehouses, Ljubljana, Slovenia.

Popović, D., & Bašić, B. D. (2009). Churn prediction model in retail banking using fuzzy C-means algorithm. Informatica, 33(2).

Raj, R., Gupta, S., Mishra, R., & Malik, M. (2024, April). Efficacy of customer churn prediction system. In 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 546-553). IEEE.

Radovanović, S., Bohanec, M., & Delibašić, B. (2023). Extracting decision models for ski injury prediction from data. *International Transactions in Operational Research*, *30*(6), 3429-3454.

Safinejad, F., Noughabi, E. A. Z., & Far, B. H. (2018). A fuzzy dynamic model for customer churn prediction in retail banking industry. Applications of Data Management and Analysis: Case Studies in Social Networks and Beyond, 85-101.

Taskin, N. (2023). Customer Churn Prediction Model In Telecommunication Sector Using Machinelearning Technique. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1774181&dswid=-3518