TWO-LAYER ARCHITECTURE OF TELCO CHURN AUTO-ML

Marin Mandić, Goran Kraljević
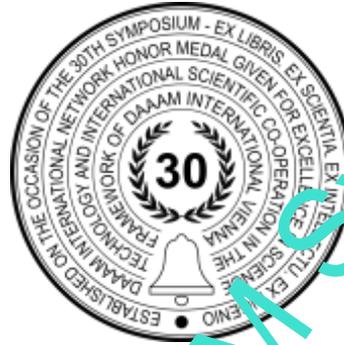
Abstract

In recent years application of machine learning has been used in many businesses and many different use cases. This led to increased need for machine learning applications. Creating machine learning applications is time consuming and requires expert knowledge of machine learning and domain expertise. This caused a development of the new research topic of Automated Machine Learning (AutoML). This research uses real telecom data to check whether the AutoML can successfully predict customer churn. In addition this research proposes the two-layer architecture of AutoML telco systems for self repairing/upgrading the model. In this work we propose the excluding the one period of implementation window, which is enabled by the use of AutoML.

Keywords: Machine learning; AutoML; Two-layer architecture of AutoML; Telco Prepaid churn prediction model

1. Introduction

Data analysis is crucial for companies to get the information they need for decision making, relationship with a customer, developing products and improving revenue and productivity. Modelling machine learning applications is needed to create advanced data analysis.

Developing good models from a dataset requires great expert knowledge and effort and much manual work. Those experts need to have domain expertise, know the process of creating machine learning applications and know how to prepare datasets. They also have to be familiar with various machine learning algorithms, they have to know how to evaluate results and how to implement those applications. In this process many different models with different hyperparameters and different attributes should be compared. That requires a lot of manual work. Also increased need for machine learning applications led to lack of experts in machine learning. All this caused the development of new research topic of Automated Machine Learning (AutoML). AutoML helps experts to automate manual repetitive task and spare time for problem analysis and deployment. AutoML helps novice data scientists and business users to create machine learning applications without machine learning expert knowledge. One more benefit of AutoML is that AutoML reduces the cost of the machine learning application.

There are various definitions of AutoML. AutoML is the process of automating the end-to-end process of applying machine learning to real-world problems [1]. AutoML is the problem of automatically (without human input) producing test set predictions for a new dataset within a fixed computational budget [2]. AutoML can lead to improved performance while saving substantial amounts of time and money, as machine learning experts are both hard to find and expensive [3].

Most AutoML tools and research papers automate the following steps of modelling machine learning applications: data preparing, model selection, hyper parameter optimization and meta-learning.

Data preparation is the most time-consuming step that usually takes up 50-70% [4] of the time of data mining projects. Therefore, it is justified to automate this manual and time-consuming work. This step includes numerous tasks: data cleaning, data normalization, attribute selection, attribute generation, binning etc. For example, Rapid Miner proposes that attributes that have correlation, high diversity, high stability and attributes with missing values should be excluded from future modelling.

Model selection and hyper-parameter optimization is most demanding and the most important part of AutoML. CASH Combined Algorithm Selection and Hyperparameter Optimization problem have been formulated by Auto-VEKA [5] in which the choice of a classification algorithm is modeled as a categorical variable, the algorithm hyper-parameters are modelled as conditional hyper-parameters. This task automatically chooses an algorithm and its hyper-parameters for a given dataset.

Machine learning tools (such as H20, DataRobot, Rapidiner) enable analysis and interpretation of created models. They provide model result through confusion matrices, various metrics (Accuracy, precision, recall, AUC, F1 etc.), various visualization and attribute importance [6]. RapidMiner even provides a model simulator to get a better insight into the model. This is especially useful for the model built by deep learning algorithm.

Comparing all possible algorithms with all possible parameters would be time-consuming and not practical.

Meta-learning [7] is concept that covers learning based on previous experience in a data-driven way. SmartML [8] is a meta learning-based framework that extracts its meta-features and searches its knowledge base for the best performing algorithm to start its optimization process.

In the past couple of years, many AutoML tools have been developed: Auto-WEKA, H20-AutoML, Auto-Sklearn, Google AutoML, DataRobot, RapidMiner, AzureML, Auto-keras etc. Comparison of many AutoML tools and algorithms was made in the following scientific papers [6] and [9].

The key contributions of this paper are:
- Introducing two-layer architecture of AutoML. Two-layer architecture enables the continuous development of challenger models which can take place of the production model and in such a way repair the production model. In this way, we propose the more complete automation of machine learning process.
- An experiment that will show that valuable results can be obtained using AutoML and real telecom data.
- Proposing excluding the implementation window from data time frame because the AutoML solution significantly reduces the time needed for collecting data and developing models.

2. Telecom churn problem

In the past couple of years, application of machine learning has been used in many businesses (banking, insurance, sales, production etc.) and many different use cases (predictive maintenance [10], segmentation, fraud detection etc.). Telecom industries use machine learning for churn prediction, fraud detection and developing products. This paper deals with the telecom churn prediction of prepaid users and it is a very important issue for telecom companies. Research up to date [11] has shown that it is five to six times more expensive to bring in a new user than to keep an existing one. It is estimated that the average customer churn in the telecom industry at 2.2% on a monthly basis [12]. Predicting customer churn opens up the possibility for telecom to try to retain such customers through retention marketing campaigns and consequently retain the revenues for the telecom. In this work we use real telecom data from telco operator from Bosnia and Herzegovina which is a small market with a low total number of customers [13]. Developing churn prediction and retention strategy can help to deal with the problem of a small market and a low number of customers.

3. Introducing Two Layer Architecture of Telco Churn AutoML

Machine learning models that are built cannot last forever because it is natural that the data is changed over time so the quality of the model fades over time. As far as we know there are no research papers about architectural design that automates self-repairing the model. In this paper, we propose two-layer architecture that introduces continuous development of challenger models. Challenger models will continuously be built based on control data and compare with the production model. If challenger model produces better evaluation result, then challenger model will be chosen to be the production model.

Lets assume that data is processed on monthly basis on the telco churn prediction problem. Every month production model generates list of users with high probability of churn. These users are offered retention marketing campaigns. To develop and evaluate challenger models control data set needs to be separated and excluded from the action of marketing campaigns. We propose that 10 percent of data is separated as a control data set. Every month, based on control data set and AutoML, challenger models are created. Those models and the production model are evaluated and if the challenger model is significantly better than production model than that model becomes a production model. Proposed two-layer architecture of telco AutoML is shown in figure 1.
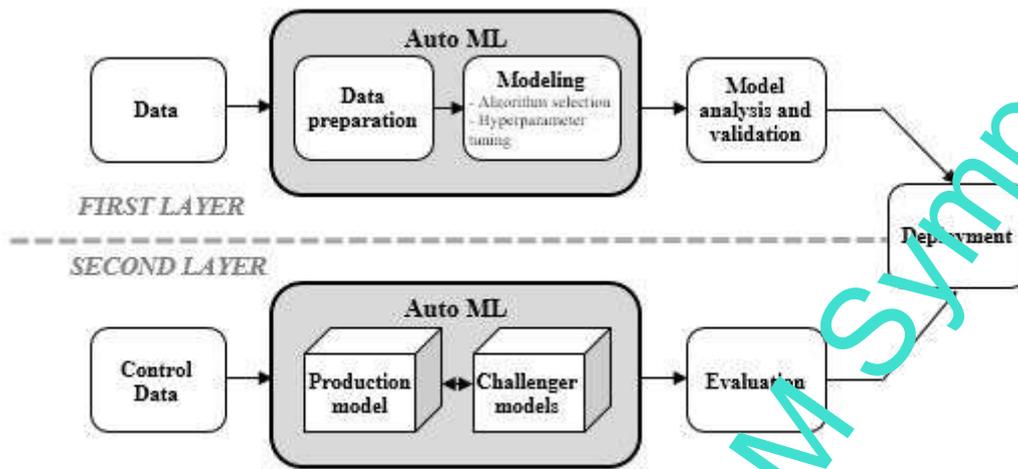
Fig. 1. Two-layer architecture

Two-layer architecture should enable further automation and easier machine learning application development.

4. AutoML for predicting telco customer churn

In this section we use AutoML for application in the telecom industry to solve the problem with the predicting customer churn. Input data used in this experiment has 206 attributes including target attribute and 56.612 records. The number of users indicated as churners is 8.396, which means there are 14.83% of the churners.

We used four months of behavioural customer data and one month of customer churn. We used four months of behavioural data because we develop trend attributes that help us to develop a more accurate model. Trend attributes have been calculated as the difference between the period that precedes churn period and an average of three preceding periods.

Usually, there is one period of implementation window [14] and in this work there is no implementation window. Reason for excluding implementation window is that complete AutoML solution significantly reduces the time needed for collecting data and developing models. Omitting this period helps to get a more accurate model because we use more recent behavioural data for predicting churn. Data time frame is shown in figure 2.

...	2019-03	2019-04	2019-05	2019-06	2019-07	...

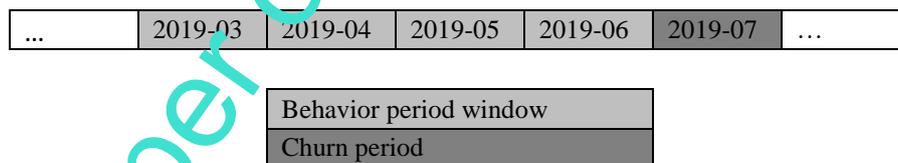| Behavior period window |
|------------------------|
| Churn period |

Fig.2. Data time frame

We used RapidMiner AutoModel to create churn prediction models. RapidMiner is a semi-automated tool that requires data scientists interaction to develop models. RapidMiner requires that data scientist import data, define the problem that needs to be solved, select target attributes, select input attributes, define the machine learning model, and analyse the result. There are also tools that are trying to automate machine learning process like Auto Weka, Hyperopt-Sklearn, Auto-Sklearn and TPOT [1].

RapidMiner AutoML proposes that 88 of total 205 attributes to be used for developing the model and the rest of them need to be filtered. We used logistic regression, deep learning and gradient boosted tree to develop a churn prediction model. Visual comparison of developed modes are shown in figure 3 using ROC curve. The result of AutoML based on real telco data is presented in tables 1, 2, 3 and 4.
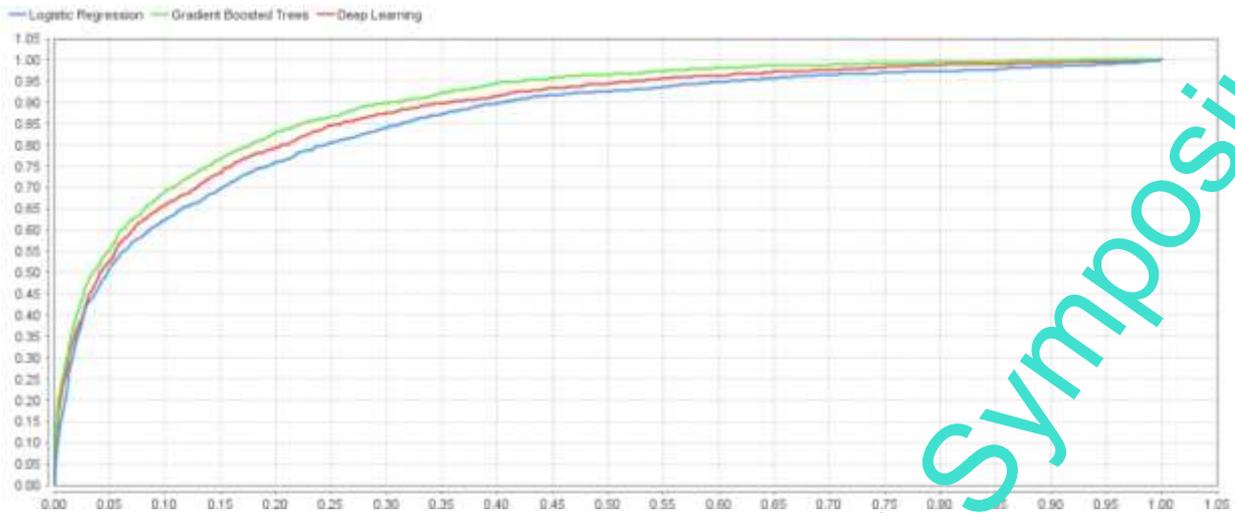
Fig.3. ROC curve comparison of the developed model

Deep learning	True non churners	True churners	Class precision
Predicted non-churners	12635	887	93.44%
Predicted churners	1155	1497	56.45%
Class recall	91.62%	62.79%	

Table 1. Deep learning confusion matrix

Logistic regression	True non churners	True churners	Class precision
Predicted non-churners	13342	1373	90.67%
Predicted churners	429	1030	70.60%
Class recall	96.88%	42.86%	

Table 2. Logistic regression confusion matrix

Gradient Boosted Tree	True non churners	True churners	Class precision
Predicted non-churners	12848	922	93.30%
Predicted churners	942	1462	60.82%
Class recall	93.17%	61.33%	

Table 3. Gradient boosted tree confusion matrix

Model	Accuracy	Precision	Recall	F measure	AUC
Deep learning	87.37	56.45	62.81	59.44	0.877
Logistic regression	88.86	70.63	42.92	53.35	0.858
Gradient Boosted Tree	88.48	60.81	61.33	61.05	0.893

Table 4. Comparing measure for developed models

From the data shown in Table 4. it is clear that the RapidMiner AutoML model measures have high values. This results provide a positive answer to the question given in the abstract, whether AutoML can successfully predict the user's churn. In addition, it can be noticed is that there are no major differences in the quality of the developed models. The deep learning model has the best recall value, slightly better than the recall of the gradient boosted tree. Logistic regression has the best precision result but recall of logistic regression is notably lower. We can conclude that the gradient boosted tree

has the highest values of F measure and AUC measure and balanced result precision and recall measure and as such is the best constructed model.

To verify proposed two-layer architecture, we used control data from the next period 2019-08 and used them to develop a challenger models. The results showed that the initially developed production model based on the gradient boosted tree still shows the best results. This is also logical because user behaviour could not change significantly after a one month. There is a limitation in the verification of the architecture because it is checked over one period of time. Future work should verify how the specified two-layer architecture behaves over a long period of time.

5. Conclusion

This paper presents a brief overview of research of Automated Machine Learning which has gained a lot of focus in the recent years. We also explain the telecom churn prediction issue because it is important for telecom companies to manage churn. An experiment was made in this work with AutoML and real telecom data which showed that satisfactory results can be obtained. Furthermore, a comparison of the performances of the three models was made and gradient boosted tree show best result. We noticed that it is possible to remove one data period using the AutoML for implementation and therefore it is possible to use more recent behavioural data. That makes it possible to get better modelling result.

The main emphasis of this paper is to propose two-layer architecture of AutoML. Two-layer architecture introduce the continuous development of challenger models which can take place of the production model and in that way repair the production model. Continuous development of challengers models enables automation of machine learning application maintenance in the production environment. Further research should verify how the specified two-layer architecture behaves over a long period of time.

6. References

[1] Chauhan, K.; Shreena, J; Thakkar, D.; Ridham D.; Bhatia, J.; Tanwar, S. & Obaidat, M. (2020). Automated Machine Learning : The New Wave of Machine Learning, Proc. Second Int. Conf. Innov. Mech. Ind. Appl., no. Icimia, pp. 205–212

[2] Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J. T.; Blum, M. & Hutter, F. (2015). Efficient and Robust Automated Machine Learning, Proc. 28th Int. Conf. Neural Inf. Process. Syst. -, vol. Volume 2, pp. 2755–2763

[3] Hutter, F.; Kotthoff, L. & Vanschoren, J. (2019). Automatic Machine Learning: Methods, Systems, Challenges, Springer International Publishing , ISBN: 978-3-030-05318-5

[4] Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.R. & Wirth, R (2000). Crisp-Dm 1.0 Step-by-step data mining guide, SPSS

[5] Thornton, C.; Hutter, F.; Hoos, H. H. & Leyton-Brown, K. (2013). Auto-WEKA : Combined Selection and Hyperparameter Optimization of Classification Algorithms, Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. data minin, pp. 847–855, ISBN: 978-1-450-32174-7, New York, USA

[6] Truong, A.; Walters, A.; Goodsitt, J.; Hines, K.; Bruss, C. B. & Farivar, R. (2019). Towards Automated Machine Learning : Evaluation and Comparison of AutoML Approaches and Tools, IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1471-1479, IEEE, DOI: 10.1109/ictai.2019.00209

[7] Vanschoren, J. (2019). Meta-Learning. In: Hutter F., Kotthoff L., Vanschoren J. (eds) Automated Machine Learning, The Springer Series on Challenges in Machine Learning. Springer, Cham. ” pp. 39–68

[8] Maher, M. & Sakr, S. (2019). SmartML : A Meta Learning-Based Framework for Automated Selection and Hyperparameter Tuning for Machine Learning Algorithms, EDBT, pp. 554–557, Lisbon, Portugal, DOI: 10.5441/002/edbt.2019.54

[9] Tuggener, L.; Amirian, M.; Rombach, K.; Lorwald, S.; Varlet, A.; Westermann, C. & Stadelmann, T. (2019). Automated Machine Learning in Practice : State of the Art and Recent Results, 6th Swiss Conference on Data Science (SDS) pp. 31–36, IEEE, Bern, Switzerland, DOI: 10.1109/SDS.2019.00-11

[10] Katona, A. & Panfilov, P. (2018). Building Predictive Maintenance Framework for Smart Environment Application Systems, Proceedings of the 29th DAAAM International Symposium, ISSN 1726-9679, ISBN 978-3-902734-20-4, B. Katalinic (Ed.), pp. 460–470, Published by DAAAM International, Vienna, Austria, DOI: 10.2507/29th.daaam.proceedings.068

[11] Lazarov, V. & Capota, M. (2007). Churn Prediction, Business Analytics Course TUM Computer Science

[12] Wie, C. & Chiu, I. (2002). Turning telecommunication call details to churn prediction: a data mining approach, Expert Systems with Applications, vol. 23, no. 2, pp. 103–112, DOI: 10.1016/S0957-4174(02)00030-1

[13] Kafol, C. & Hocevar, M. (2016). Consolidation on Small Telecommunication Market, Chapter 09 in DAAAM International Scientific Book 2016, B. Katalinic (Ed.), pp. 93-102, Published by DAAAM International, ISBN 978-3-902734-09-9, ISSN 1726-9687, Vienna, Austria

[14] Lu, N.; Lin, H.; Lu, J. & Zhang, G. (2014). A Customer Churn Prediction Model in Telecom Industry Using Boosting, IEEE Transactions on Industrial Informatics vol. 10. c, pp. 1659-1665, DOI:

10.1109/TII.2012.2224355
[15] Nagarajah, T. & Guhanathan, P. (2019). A Review on Automated Machine Learning (AutoML) Systems, IEEE 5th International Conference for Convergence in Technology, pp. 1–6, IEEE, Bombay, India, DOI: 10.1109/I2CT45611.2019.9033810