Gradient Boosting Decision Trees for Patient Survival Prediction

Sedigheh Eslami
Data4Life
Potsdam, Germany
sedigheh.eslami@data4life.care

Evgeniya Anikina
Data4Life
Potsdam, Germany
evgeniya.anikina@gmail.com

Ariane Morassi Sasso
Hasso Plattner Institute
Potsdam, Germany
ariane.morassi-sasso@hpi.de

ABSTRACT

Correct survival prediction highly affects the decision making process for critical patients. Learning from real-world clinical data can help healthcare systems predict the survival rate more accurately. However, such data usually suffers from an imbalance in the survival cases distribution. In this work, we evaluated two gradient boosting decision tree models, LightGBM and Catboost. Based on our experiments, Catboost achieved a better test AUC score (0.91) in the task of binary classification with imbalanced data. Here we detail the work of the team *AcES2Predict* in the WiDS 2020.

KEYWORDS

Survival Prediction, Gradient Boosting Trees, Catboost, LightGBM

1 INTRODUCTION

Predicting the severity of patients' conditions in Intensive Care Units (ICUs) highly affects the process of decision making in healthcare systems. Many solutions have been proposed to evaluate the effectiveness of clinical treatments and interventions in ICUs based on severity scores, such as the Acute Physiology and Chronic Health Evaluation (APACHE) and the Simplified Acute Physiology Score (SAPS) [6, 7]. However, besides requiring regional customization and constant updates, they are mainly evaluated using only data from the United States [12]. Additionally, these scoring systems use parametric linear models, e.g. logistic regression, which are not always effective due to their lack of flexibility [9]. Strategies based on non-parametric models such as *Super Learner* and EMPICU have already outperformed these state-of-the-art scores by using ensemble decision trees algorithms—specifically Random Forests (RF) and Bayesian Additive Regression Trees (BART) [2, 9]. In this work, we then evaluate the performance of two recently developed Gradient Boosting Decision Trees (GBDT) models on the international MIT's GOSSIS ICU visits dataset. To the best of our knowledge, this is the first work using data from ICUs to compare GBDT for patient survival prediction in an imbalanced setting.

2 METHODS

Real-world clinical data often suffers from the imbalanced class distribution problem. In fact, the initial exploration of the data released in the WiDS 2020 challenge also illustrated the imbalance in the mortality class distribution. Our approach to overcome this problem was to use ensemble models, specifically gradient boosting techniques. We performed our experiments with LightGBM and Catboost [5, 10] due to their success in previous work [1, 8, 14].

Permission
AcES2Predict, WiDS 2020

3 EXPERIMENTS AND INSIGHTS

Dataset. We used WiDS 2020 challenge dataset in our experiments. This dataset contains more than 180 demographic and clinical features for more than 90K unique patients from different countries[1]. In the training subset, more than 30% of the values were missing. We decided to drop {*encounter_id, patient_id, hospital_id* and *icu_id*} columns as they were randomly assigned and uninformative. The ratio of positive (death) to negative (survived) cases is about 1:10.
Results. We executed LightGBM and Catboost with log-loss [13] as the evaluation metric in training, with no prior data resampling[2]. Catboost is capable of class-weighted (cost sensitive) learning in order to solve the class imbalance problem. We set the class weight to [1, 10], which is the ratio of dead vs. survived classes. LightGBM and Catboost both have built-in mechanisms for completing the missing values in the data. Therefore, we did no prior handling of the missing values. In both models, we performed a 6-fold Cross Validation (CV). The corresponding results for train loss and AUC scores [3] are illustrated in Table 1.

Table 1: LightGBM and Catboost results on WiDS 2020

	Train Loss	Average CV AUC	Test AUC
LightGBM	0.1	0.67	0.88
Catboost	0.29	0.81	**0.91**

Insights. Our experiments indicate that Catboost outperforms LightGBM in the task of survival prediction. We hypothesized that this might be due to Catboost encoding categorical features better while being less biased [4]. Secondly, with Catboost we achieve a cost-sensitive classifier with respect to the class weights which refuses to get biased towards the majority class as a result. Furthermore, it is worth noticing that our LightGBM is over-fitting the training data with a loss of 0.1.

4 CONCLUSION AND FUTURE WORK

Catboost achieved the best AUC score of 0.91 on the imabalanced WiDS 2020 dataset. However, Purushotham et al. shows that ensemble strategies based on Deep Learning (DL) algorithms can outperform both the traditional scores and also *Super Learner*. Thus, in our future work, we would like to investigate Catboost's performance against traditional scores (e. g. SAPS), *Super Learner* and ensemble DL algorithms. We should also evaluate other metrics such as the F measure or the Precision-Recall Area Under the Curve.

[1] More information at https://www.kaggle.com/c/widsdatathon2020/data
[2] Source code can be found at https://github.com/sarahESL/WiDS2020

REFERENCES

[1] Essam Al Daoud. [n.d.]. Comparison between XGBoost, LightGBM and CatBoost Using a Home Credit Dataset. ([n. d.]).

[2] Aya Awad, Mohamed Bader-El-Den, James McNicholas, and Jim Briggs. 2017. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *International Journal of Medical Informatics* 108 (dec 2017), 185–195. https://doi.org/10.1016/j.ijmedinf.2017.10.002

[3] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159.

[4] Yunru Huang, Mengyu Li, and Yun Wu. [n.d.]. KKBox's Music Recommendation. ([n. d.]).

[5] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie Yan Liu. 2017. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, Vol. 2017-Decem. NIPS 2017, Long Beachh, CA, USA, 3147–3155. https://github.com/Microsoft/LightGBM

[6] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman. 1985. APACHE II: A severity of disease classification system. *Critical Care Medicine* 13, 10 (1985), 818–829. https://doi.org/10.1097/00003246-198510000-00009

[7] J. R. Le Gall. 1993. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA: The Journal of the American Medical Association* 270, 24 (dec 1993), 2957–2963. https://doi.org/10.1001/jama.270.24.2957

[8] Marco Mamprin, Jo M Zelis, Pim A L Tonino, Svitlana Zinger, and Peter H. N. de With. 2020. Gradient Boosting on Decision Trees for Mortality Prediction in Transcatheter Aortic Valve Implantation. (2020). arXiv:2001.02431 http://arxiv.org/abs/2001.02431

[9] Romain Pirracchio, Maya L. Petersen, Marco Carone, Matthieu Resche Rigon, Sylvie Chevret, and Mark J. van der LAAN. 2015. Mortality prediction in the ICU: can we do better? Results from the Super ICU Learner Algorithm (SICULA) project, a population- based study. *Lancet Respir Med* 3, 1 (2015), 42–52. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations arXiv:NIHMS150003

[10] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. Catboost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, Vol. 2018-Decem. Neural information processing systems foundation, Pre-print, 6638–6648. arXiv:1706.09516 http://arxiv.org/abs/1706.09516

[11] Sanjay Purushotham, Chuizheng Meng, Zhengping Che, and Yan Liu. 2018. Benchmarking deep learning models on large healthcare datasets. *Journal of Biomedical Informatics* 83 (2018), 112–134. https://doi.org/10.1016/j.jbi.2018.04.007

[12] Jorge I.F. Salluh and Márcio Soares. 2014. ICU severity of illness scores: APACHE, SAPS and MPM. , 557–565 pages. https://doi.org/10.1097/MCC.0000000000000135

[13] Vladimir Vovk. 2015. The fundamental nature of the log loss function. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9300. Springer Verlag, 307–318. https://doi.org/10.1007/978-3-319-23534-9_20 arXiv:1502.06254

[14] Yuan Xie, Bin Jiang, Enhao Gong, Ying Li, Guangming Zhu, Patrik Michel, Max Wintermark, and Greg Zaharchuk. 2019. Use of Gradient Boosting Machine Learning to Predict Patient Outcome in Acute Ischemic Stroke on the Basis of Imaging, Demographic, and Clinical Information. *American Journal of Roentgenology* 212, 1 (2019), 44–51.