Building an injury prediction model using football dataset with unbalanced classes

Aleksandra Sadurska[1] and Tomasz Piłka[2]

[1] Faculty of Mathematics and Computer Science
Adam Mickiewicz University in Poznań, Poland
alesad3@st.amu.edu.pl

[2] Faculty of Mathematics and Computer Science
Adam Mickiewicz University in Poznań, Poland
tomasz.pilka@amu.edu.pl

Abstract: Every year we observe growing expectations towards the performance of professional football players. Increasing overall player load, leads to higher risk of occurrence of an injury. Injuries have a significant impact not only on football team performance but also on club budget. The objective of this research is to present results of injury prediction models implemented with various machine learning algorithms and sampling methods. The dataset and results were compared with a similar study, to comprehend whether the collected dataset is sufficient to predict an injury with certain precision. Combination of XGBoost algorithm and ROSE sampling method gave the best prediction results (F-Measure: 0.40, Precision: 0.25, Recall: 1.0).

Keywords: Football; Injury Prediction; Sports Analytics; Unbalanced Data

1 Introduction

1.1 Why is injury prediction important?

Every year we observe growing expectations towards the performance of professional football players. Increasing the number of high intensity runs, accelerations, decelerations and in the result overall player load, leads to higher risk of occurrence of an injury [1]. Injuries have a significant impact not only on football team performance but also on club budget. According to research made in English Premier League a team loses an average of £45 million due to injury-related decrement in performance per season [2].

1.2 Research objectives

The objective of this research is to present results of injury prediction models implemented with various machine learning algorithms and sampling methods. This kind of experiment has been done before, for example by Bruno Goncalo Pires Martins in his study [3]. However, in the aforementioned research, models were trained on dataset collected from another league and with additional parameters like heart rate or rate of perceived exertion. Later in this article, I will compare the dataset and results used in both experiments, to comprehend whether the collected dataset is sufficient to predict an injury with certain precision.

1.3 What kind of injury can we predict?

Football is a high intensity sport with many interactions between players, like headings or tackles. Many of them may cause injuries that are clearly impossible to predict. On the other hand, just as common are non-contact injuries, usually resulting from over- or undertraining. They are caused by too high or too low player load within a certain timeframe, which depends on the applied training periodization model.

2 Dataset

2.1 Data characteristics

The data were collected during pre-season and the first round in the professional football league, among 31 players with an average age of 25. The data were collected using Catapult wearable global positioning trackers, both during training and match activities. Dataset used in this research contains information about 2733 events, described by the following parameters: High Speed Running, Maximum Velocity, Running, Velocity Band Total Effort Count, Sprint, Total Player Load, Field Time, Player Load Per Minute, Accelerations, Decelerations.

The dataset consists only of external load parameters, whereas in the study [3], the parameters of internal load, such as heart rate exertion, energy expenditure or rate of perceived exertion were also taken into account.

2.2 Unbalanced classes problem

Characteristics of this study are associated with unbalanced classes problem. From over 2700 events in the dataset, only 28 of them are labeled as an injury, so the ratio between classes is about 99:1. Due to uneven classes distribution, after splitting dataset into train and test set, following oversampling methods have been applied to trainset to balance both classes sample size: Synthetic Minority Oversampling Technique, Random Over-Sampling Example and Adaptive Synthetic [3].

3 Machine learning

3.1 Implemented machine learning algorithms

Machine learning methods of different characteristics and complexity were compared. The models were implemented using the following algorithms: Naïve Bayes, Support Vectors Machine, Random Forest, AdaBoost, XGBoost [7, 8].

3.2 Model evaluation

The most popular evaluation metric to assess the performance of a machine learning model is accuracy [4], which can be considered as the probability of identifying the right class of an observation. Nevertheless, on an unbalanced dataset accuracy can easily achieve 99% without correctly classifying any of the target examples. Therefore, it is not an appropriate method of evaluating an injury prediction model.

In this case it is more accurate to use precision, recall and F1-measure metrics [5]. From a perspective of injury prediction, the exclusion of false positives is most expected, because potentially stopping a player in a particular match with no reason not only can cost the team a points, but also trust for predictive analysis.

3.3 Results

Results for each model and balancing method are presented in Table 1. Each combination is described by F-measure, precision and recall evaluation metrics. Prediction was done on a test collection of 543 non-injury and 4 injury events .

Table 1. Results for each balancing method and model

Balancing	Model	F-Measure	Precision	Recall
ROSE	Naïve Bayes	0.02	1.0	0.01
ROSE	SVM	0.04	0.5	0.02
ROSE	Random Forest	0.10	0.5	0.06
ROSE	AdaBoost	0.15	0.75	0.08
ROSE	**XGBoost**	**0.40**	**0.25**	**1.0**
ADASYN	Naïve Bayes	0.02	0.75	0.01
ADASYN	SVM	0.04	0.5	0.02
ADASYN	Random Forest	0.04	0.25	0.02
ADASYN	AdaBoost	0.05	0.5	0.03
ADASYN	XGBoost	0.15	0.25	0.11
SMOTE	Naïve Bayes	0.02	0.75	0.1
SMOTE	SVM	0.04	0.5	0.02
SMOTE	Random Forest	0.04	0.25	0.02
SMOTE	AdaBoost	0.05	0.5	0.03
SMOTE	XGBoost	0.17	0.25	0.12

Combination of XGBoost algorithm and ROSE sampling method gave the best prediction results (F-Measure: 0.40, Precision: 0.25, Recall: 1.0). There is still room for improvement, but what is important, is that there are no non-injury events classified as injury, so no false positive results. In the confusion matrix below (Fig. 1), distribution of correctly and incorrectly classified items is shown.

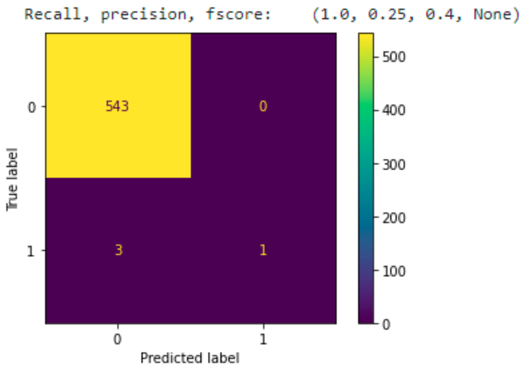

Fig 1. Confusion matrix for XGBoost algorithm and ROSE sampling method

In study [3], despite differences in dataset structure, XGBoost combined with the ROSE sampling method also gave the best prediction results (F-Measure: 0.22, Precision: 0.13, Recall: 0.67).

4 Conclusions

The results of both studies give a good perspective for further work, but in order to consider the prediction model ready for use in practice, its effectiveness should be improved. According to research conducted so far there are plenty of possibilities to achieve this goal. In the study [6], authors presented a multi-dimensional model to predict whether a player will get injured or not. Mentioned model uses lots more and also more complex parameters, like Acute:Chronic Workload Ratio, Exponential Weighted Moving Average or previous injury. Extending the dataset with additional features, like external factors related with training or internal factors, such as heart rate during activity could improve model effectiveness. Nevertheless, what should be pointed out is that injury prediction is not a fully explored problem yet. Due to the complexity of the problem, there is no single confirmed way to improve the model results.

References

1. Bengtsson, H, Ekstrand, J, Hägglund, M. Muscle injury rates in professional football increase with fixture congestion: an 11-year follow-up of the UEFA Champions League injury study. Br J Sports Med. 2013;47(12):743–747.
2. Eliakim E, Morgulev E, Lidor R, et al Estimation of injury costs: financial damage of English Premier League teams' underachievement due to injuries; BMJ Open Sport & Exercise Medicine 2020;6:e000675. doi: 10.1136/bmjsem-2019-000675
3. Martins, B. Predicting the risk of injury of professional football players with machine learning.
4. Provost, F., Fawcett, T., Kohavi, R. The Case Against Accuracy Estimation for Comparing Induction Algorithms. Proceedings of the Fifteenth International Conference on Machine Learning, 445–453. https://doi.org/10.1.1.49.5218

5. Hossin M, Md Nasir S. A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process 5.2 (2015): 1.
6. Rossi A., Pappalardo L., Cintia P., Iaia F., Fernández J., Medina D. Effective injury forecasting in soccer with GPS training data and machine learning. PLOS ONE. 13. e0201264. 10.1371/journal.pone.0201264.
7. Soofi A. A., Awan A. Classification Techniques in Machine Learning: Applications and Issues, 459-465. https://doi.org/10.6000/1927-5129.2017.13.76
8. Sagi O, Rokach L. Ensemble learning: A survey. WIREs Data Mining Knowl Discov. 2018; 8:e1249. https://doi.org/10.1002/widm.1249