Journal of Physical Education and Sport ®(JPES),Vol 21(Supplement issue 2), Art 123 pp 990–998, Apr. 2021
online ISSN: 2247 - 806X; p-ISSN: 2247 – 8051; ISSN - L = 2247 - 8051 © JPES

Original Article

Optimization of the structure of the speedway team in the presence of regulation and financial constraints –application of machine learning and integer programming to the Polish league case

SYLWESTER BEJGER
The Faculty of Economic Sciences and Management, Nicolaus Copernicus University, Toruń, POLAND

Published online: April 30, 2021
(Accepted for publication April 15, 2021)
DOI:10.7752/jpes.2021.s2123

Abstract:
In a difficult macroeconomic environment, managing a sport club and a sport team is not an easy task. It is a special challenge due to financial, demographic and organizational constraints arising and changing permanently. One of the main tasks faced by managers in the team sports is to build the optimal composition of the team, taking into account the measures that define the sports goal. The article discusses the problem of team structure optimization with the use of statistical data on the example of the Polish highest level speedway league, PGE Ekstraliga. Taking into account the demographic predictor (age of the riders) and the nationality of the riders, an analysis of the importance and impact of these predictors on the rider's performance was carried out. Using machine learning methods, the characteristic profiles of the riders, differing in effectiveness, were also indicated. The method used allows for local interpretation of prediction, as well. As a final step of the research an optimization mechanisms of team building was developed. The study proposes an optimization model supporting the decision making process of creation of the structure of the team registered for the meeting. Integer programming optimization mechanisms were constructed in concordance with possible financial and organizational constraints which could be faced by team managers. Models' validation with the use of numerical parameters' values showed their internal consistency and provided interesting decision's indications. Both machine learning method and optimization models can be develop further to cover more reach data sets and more managerial constraints or even could be tuned to be applied in the other team sports.
Key Words*:***sport team management, motorcycle speedway, optimization of managerial decisions, machine learning**

Introduction

Motorcycle speedway (or speedway in short, when misunderstanding is impossible) was created in Australia at the beginning of the 1920s (May, 1978). The competition in speedway takes place on a loose surface track in the shape similar to an ellipse. The track length is variable and on average ranges from 300 to 400 meters. The basic unit of competition is a heat. In a typical meeting the heat is run by four riders, scoring 3, 2, 1 and 0 points, depending on the place at the finish line. The heat takes four laps of the track.

Speedway is an individual sport in which trained skill of the rider plays a central role (May, 1978). There are a lot of individual tournaments, like country's Championships or Speedway Grand Prix where individual effort of the rider is on the top. However, taking into account popularity and social, economic and sporting role the national speedway leagues are undoubtedly the most important elements of speedway sport. Poland is a good example here. In Poland three levels of speedway league exist, namely: PGE Ekstraliga (the highest level league), e-Winner 1 Liga Żużlowa (speedway first league) and 2 Liga Żużlowa (speedway second league). In those leagues participate 23 teams in total (as assigned for season 2021). Especially important is PGE Ekstraliga, considered as one of the most important speedway leagues in a world. In a typical season about 170 riders is assigned to Ekstraliga teams. In terms of popularity, the meetings of the PGE Ekstraliga alone gathered around 700,000 fans in the 2019 season (PGE statistics, 2021). A speedway sport, highly connected with technical and logistics infrastructure, is a costly one both for the riders and the clubs. For example, top rider's of PGE Ekstraliga expenses per season amounts about 800.000 PLN (see Koerber, 2020 for detailed calculations). In the other hand clubs have to pay riders for points collected during meetings and for contract signed per season, have to cover all of the expenses connected with track maintenance, club operations and meetings' organisation. A budget of a speedway club participating in PGE Ekstraliga could be about 5.000.000 PLN per season (Koerber, 2020). In the light of the above-mentioned facts, it should be considered that it is the team games that determine the survival of the speedway sport discipline. Contrary to individual tournaments, the success of the team is the sum of not only individual efforts of the riders, but also depends on the proper club and

Corresponding Author: SYLWESTER BEJGER, E-mail: sylwester.bejger@umk.pl

team management and effective strategy of team building. Improper management may lead to financial difficulties, which in turn might result in bankruptcy of the club (Lis, Tomanek, 2020).

Professional management of sport organization combines skills related to planning, organizing directing, controlling, budgeting, leading and evaluating within the context of an organization or department whose primary product or services is related to sport and/or physical activity (DeSensi, et al., 1990). Sports organizations which have as object of activity professional sport are particularly concerned to ensure all conditions for winning sports competitions, namely achieving high performance.

To achieve this objective, the management within these organizations should be focused on sports team management, in other words, should focus on the human factor that decides the outcome of sports competitions. Therefore, it is necessary to ensure an efficient management of human resources at the club or sports team level. The managerial tasks involved in that activities include setting the primary objectives and **making a decisions** on team's structure (Aćimović, et al., 2013).

Decision making process could be fully qualitative or supported by quantitative techniques. Digital transformation enables data-driven methods, as Machine Learning to be used in sport management. These methods are very effective in a sport's data exploration (Fujii, 2021) and a sport results' prediction problems (Bunkera, Susnjak, 2019), but only their combination with the proper optimization methods provides the right decision-making tool.

In the paper author concentrate on the problem of optimal decisions connected with the team structure. When managerial decision connected with operation of the club are highly complicated and, in most cases, made in environment which is publically unknown (trade secrets), the team building decisions could be analysed and optimized on a basis of statistical data. In the case of team building, it is also relatively easy to distinguish significant external parameters that limit decision set. In the case of the Polish league, these are: financial budget restrictions and competition regulations. Competition regulations has direct impact on a team's structure in two most important, variable factors – age of the riders and nationality of the riders in team. Focusing on the Polish highest level league, the PGE Ekstraliga, the main goals of the paper are twofold. First of them is analysis of impact of age and nationality structure of an average team on efficiency of team, measured by points collected by the riders. Analysis should discover relative importance of each feature, as well. Machine learning algorithm of random forest was utilised in that task. The second aim is building an optimization model for a team structure taking into consideration new rules of competition, valid for season 2021. In this stage a numerical experiment has been done.

Material & methods

Description of data. Time period of the research covers seasons 2015 – 2020 of speedway PGE Ekstraliga in Poland. Author used data[1] collected by the portal http://gurustats.pl. Source data set contain 5635 records. Object of each record is a rider's performance in a single meeting. Fields of the record contain following features: date of the meeting, points collected, age of the rider, nationality tag (polish/other). The original features were transformed accordingly during examination. Description of these transformations is in the Results section.

The following research methods have been used to achieve the objectives set: subject literature analysis, machine learning and optimization.

For the preliminary examination of data basic statistical methods have been used. Author use statistical visualization (histograms, scatter plots and a kernel density estimate (KDE) plot) and descriptive statistics.

In the scope of a goal number one a Random Forest Regression algorithm was utilised. Random Forest (Ho, 1995; Breiman, 2001) is an ensemble (or forest) of decision trees grown from a randomized variant of the tree induction algorithm. Building block of Random Forest ensemble is a weak learner of a regression tree, (Breiman et al., 1984; Quinlan, 1993).

The prediction function of a tree is defined as:

$$f(x) = \sum_{m=1}^{M} c_m I(x, R_m), \qquad (1)$$

where: M is the number of leaves in a tree, R_m is a region in the feature space (corresponding to leaf m), c_m is a constants corresponding to region m, I is the indicator function (returning 1 if $x \in R_m$, 0 otherwise). The value of c_m is determined in the training phase of the tree, which in case of regression trees corresponds to the mean of the response variables of samples that belong to region R_m. For regression one partition predictor space in order to find set of regions R that minimize the RSS, given by:

$$\sum_{m=1}^{M} \sum_{i \in R_m} (y_i - \hat{y}_{R_m})^2, \qquad (2)$$

where: $\hat{y}_{R_m}$ is is the mean response for the training observations within the m-th region.

The Random Forest prediction is the unweighted average over the set of K trees:

$$F(x) = \frac{1}{K} \sum_{i}^{K} f_k(x) \qquad (3)$$

Random Forest Regression was a method of choice because it has important characteristics which make it attractive:

[1] Data available commercially on request.

- As all tree-based methods it's non-parametric. Can model arbitrarily complex (nonlinear) relations between predictors and target, without any a priori assumption,
- Can handle heterogeneous data (ordered or categorical variables, or a mix of both) without preprocessing,
- Permits ranking of the relative significance of predictor variables, through variable importance metrics VIMs; (Biau, Scornet, 2016),

VIM calculation is based on observation that features used at the top of the tree contribute to the final prediction decision of a larger fraction of the input samples. The expected fraction of the samples they contribute to can thus be used as an estimate of the relative importance of the features.
In a stage number one author exploited RFR model to obtain information on average values of predicted number of points for training set and VIM values rather than to obtain good generalisation on a test set.

At the second stage of the research a simple optimization model (in two variants) has been proposed. The model used is of integer programming class (Hillier, Lieberman; 2001). Its structure is presented in a Results section.

Results

The study started from visual exploration[2] of empirical distributions of three features which are subject of the research – *Pts* – number of points per meeting per rider (target variable), *Age* – age of a rider at the date of a meeting (in years), *Pol/Int* – categorical feature, value = y for Polish rider, value = n for other nationality. Both *Age* and *Pol/Int* are predictors. Figures 1 and 2 depicted empirical distribution of *Age* and *Pol/Int*. One can observe clustering of age distribution around few centroids (KDE show three or more peaks). There is a change in a shape of distribution from the 2018 season up. In a case of nationality, one can see greater share of foreign riders in the meetings in the seasons 2019 and 2020 then in the previous years.

Fig. 1. Histograms and KDE – Age

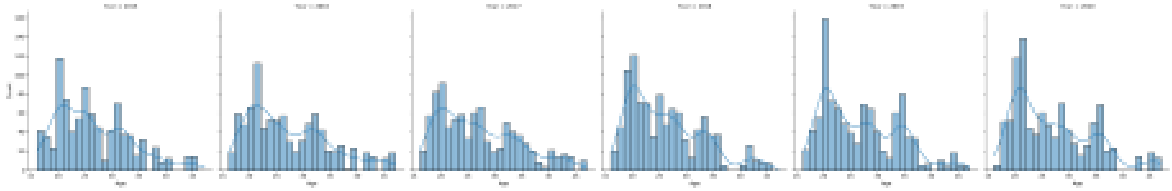
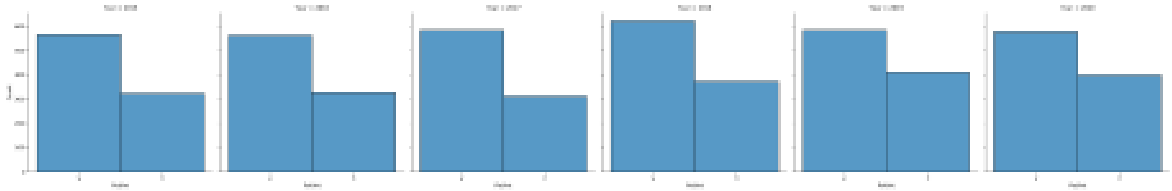


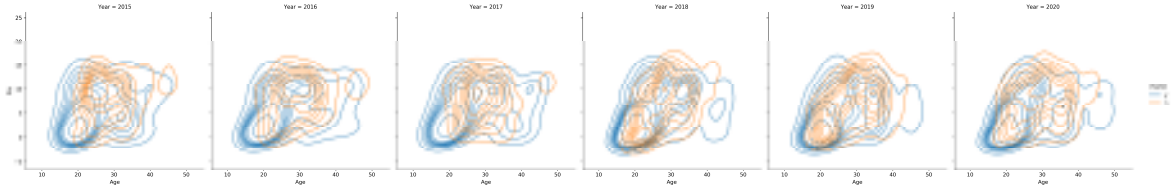
Fig. 2. Histograms – Pol/Int



In the Figures 3 and 4 visual analysis of relation between target variable and predictors is presented. The most important observation is different efficiency of riders by nationality and by age group. The KDE by season show some structural changes in distributions, as well. There exists obvious difference in efficiency between younger and older riders.

Fig. 3. Scatter plot – Pts vs. Age and Pol/Int

Fig. 4. KDE – Pts, Age and Pol/Int



[2] All of the Tables and Figures in a paper present author's own results.

To asses differences in target variable more precisely, *Age* variable was transformed into two categorical variables, one for rider younger than 25 years and second for rider older than 25 years. Using these two variables for grouping some descriptive statistics has been calculated. The results shows Table 1.

Table 1. Descriptive statistics – Pts by age and nationality

season	nationality	rider group	number of heats	mean	median	std. dev.	coefficient of variation
2015	Other	Older	238	8.67	9	3.70	43%
		Younger	86	7.33	8	4.67	64%
	Polish	Older	222	6.62	6	4.11	62%
		Younger	343	4.62	3	4.10	89%
2016	Other	Older	258	9.03	9	3.48	39%
		Younger	70	5.24	5	3.87	74%
	Polish	Older	201	7.26	8	3.78	52%
		Younger	361	4.39	3	3.85	88%
2017	Other	Older	250	8.15	8	3.75	46%
		Younger	63	6.62	7	3.92	59%
	Polish	Older	233	7.27	8	3.76	52%
		Younger	349	4.42	3	3.79	86%
2018	Other	Older	262	8.57	9	4.18	49%
		Younger	110	3.56	3	3.64	102%
	Polish	Older	241	7.37	8	3.74	51%
		Younger	382	3.51	2	3.91	111%
2019	Other	Older	279	8.52	9	4.03	47%
		Younger	132	4.52	5	3.58	79%
	Polish	Older	225	6.57	6	3.67	56%
		Younger	357	3.65	2	4.11	113%
2020	Other	Older	286	8.33	9	4.10	49%
		Younger	111	5.95	6	4.05	68%
	Polish	Older	230	6.85	7	4.16	61%
		Younger	346	2.86	2	3.19	111%

Results reported in Table 1 confirms fairly stable (expect season 2020) patterns in efficiency of the riders. On average, Younger Polish riders performed the worst, next were foreign Younger riders. The second best group was Older Polish riders, and the best was foreign Older riders. This ranking is somehow obvious but stability of performance in groups can be surprising. As proxy measure of shape's stability coefficient of variation has been used. That measure shows high instability of a performance of Younger Polish riders and more stable shape of foreign riders than the Polish riders in an Older group. To summarise, shaping a structure of a teams, taking into account point's efficiency could be very complicated task, even for two predictors.

In a first stage of an examination the more careful study of influence of age and nationality on teams' structure and efficiency has been done. To extract teams' creation rules and assess impact of the predictors on efficiency (in terms of points scored) author used machine learning algorithm of Random Regression Forest (RFR in short).

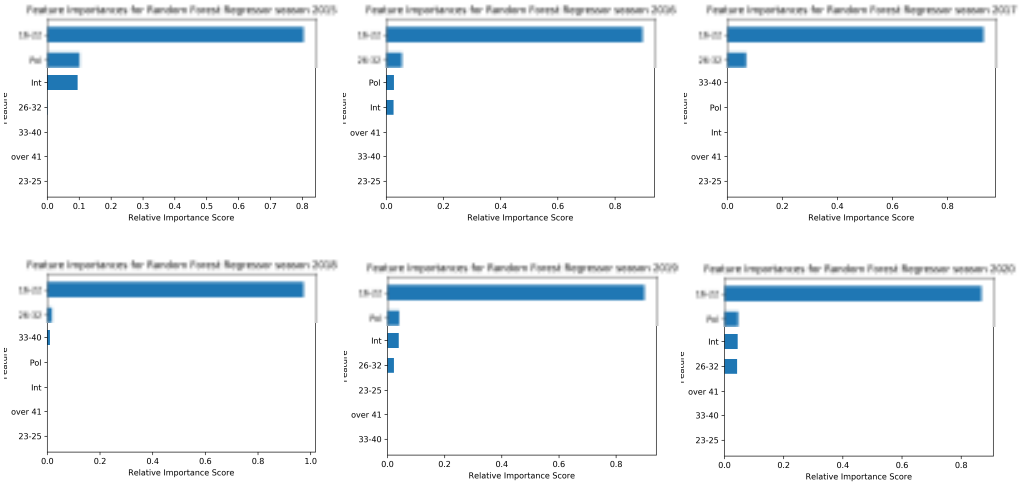
The purpose of estimation of RFR models were twofold: to evaluate the influence of the predictors on the average number of points scored and estimate prediction of average number of points for possible combinations of age and nationality. To make the estimation more informative, the *Age* variable was transformed into 5 binary variables, each one representing an age bin. Age bins was chosen on a basis on histograms presented in Fig. 1. Feature *Pol/Int* was binary encoded, either (into *Pol* feature, value = 1 mean Polish nationality and *Int* feature, value = 1 means other than Polish nationality).

The whole data set was divided into subsamples encompassing one season. For each subsample a RFR model has been estimated (hyperparametrs tuned by a greed search). Training and test set were in proportion 0.9/0.1 in each model. The set of variables included: *Pts* – target, *Pol, Int, 16-22, 23-25, 26-32, 33-40, over 41* – predictors. Names of predictors (expect *Pol*, *Int*) represent age bins.

In the main phase of a stage one six RFR models were estimated. Next the VIM measures were calculated and plotted.

Fig. 5 depicts the results.

Fig. 5. Relative importance of predictors by seasons



In better understanding of Fig. 5 helps Fig. 6 below.

Fig. 6. Local interpretation of two predicted values

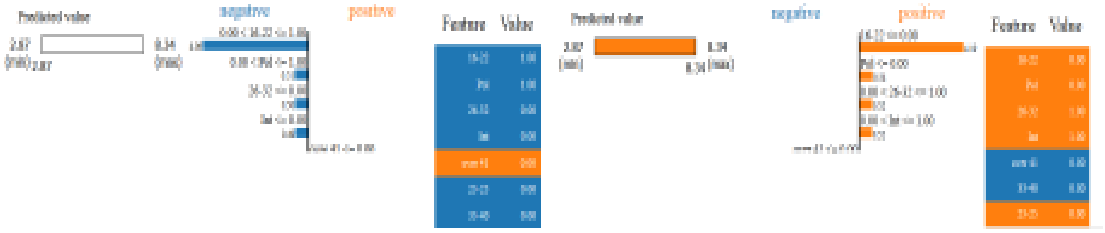


Fig. 6 shows local interpretation (using LIME, see Ribeiro et al. 2016, Bejger, Elster 2020) of prediction of average number of points collected in a season 2020 by two riders: rider number 1 (left panel), who is Polish and of an age between 16 and 22 years, rider number 2 (right panel), who is not Polish and his age is between 26 and 32 years. As one can see, feature *16-22* has the greatest influence on predicted value of point in both cases, but in a first case it is **negative influence of being** in that age group, while in a second case it is a **positive influence of not being** in that age bin. In this context, the importances showed in Fig. 5 mean the influence (negative or positive) of a given predictor on the determination of the target value. It is clear, that the greatest influence on target over all seasons had variable *16-22* but importnaces of other predictor differed significantly. In the season 2017 and 2018 nationality variables did not count very much on prediction estimation, in a contrary to the seasons 2015, 2019 and 2020 where nationality plays important role.

The RFR estimation leads to very interesting predictions of mean number of points collected by group of riders. As author used only binary variables describing two dichotomic features (age and nationality), RFR estimation allows not only for predictions but either for clustering those predictions into distinct combination of features, which can be used further in optimization phase. Table 2 explains this statement. To focus on particular season, it shows estimates for the year 2020[3].

Table 2. Prediction for a season 2020 (bin based on histograms)

	feature						
predicted mean number of points	**Pol**	**Int**	**16-22**	**23-25**	**26-32**	**33-40**	**over 41**
2.874	1	1	1	0	0	0	0
6.230	1	0	0	1	0	1	1
6.921	1	0	0	0	1	0	0
7.646	0	1	0	1	0	1	1
8.337	0	1	0	0	1	0	0

From Table 2 one can learn that Polish young rider (predictors *Pol* =1 and *16-22* = 1) will bring 2.874 points by season. In the other hand, foreign rider in the age of 35 years will score 7.646 points. These values can be a good approximation of an average meeting's value of points, either. Interestingly, Polish riders in the age

[3] Estimates for other seasons available for request from author.

bins: 23-25, 33-40 and over 41 score the same number of points, on average. It means that such riders could be used interchangeably.

Part two of the study contains proposition of an optimization model which could help determine demographic and nationality structure of the team in concordance with regulation's and financial constraints.

Author focus on season 2020 as a base season for RFR estimation, but correct the age bin according to the new "Polish team championship of PGE Ekstraliga Regulation" (Regulation, 2021) which is valid for season 2021. The most important constrains implied by the Regulation are described further in a text. The direct consequence of Regulation's rules is a new bin structure for an age feature which implies new set of categorical variables connected with.

The RFR estimation for season 2020 was repeated with these new features. VIM values and predictions for this estimation contain Fig. and Table 3.

Fig. 7 Relative importance of predictors for season 2020

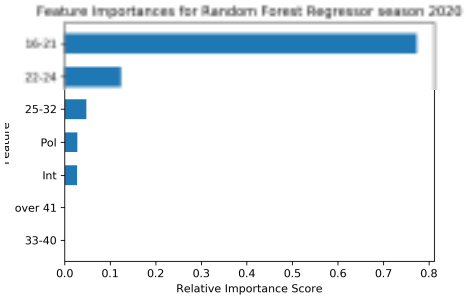


Table 3. Prediction for a season 2020 (bin's structure based on Regulation)

	feature						
predicted mean number of points	**Pol**	**Int**	**16-21**	**22-24**	**25-32**	**33-40**	**over 41**
2.232	1	0	1	0	0	0	0
2.281	0	1	1	0	0	0	0
5.132	1	0	0	1	0	0	0
5.746	0	1	0	1	0	0	0
6.710	1	0	0	0	0	1	1
7.226	1	0	0	0	1	0	0
7.324	0	1	0	0	0	1	1
7.817	0	1	0	0	1	0	0

The next step was a construction of an optimization model. The main modelling assumption were:

- Model should solve a decision problem of finding optimal speedway team's structure, where team is considered as a team registered for meeting in the PGE Ekstraliga,
- Model should base on RFR estimation of the mean values of point predicted for riders,
- Model should be coherent with 2021 Regulation,
- Model should be constrained by possible financial (budget) constraint.

The most important rules, implied by Regulation for season 2021 an connected with a team building are as follow:

a) the team registered for a meeting contains 8 riders at most (7 in the basic team plus one a reserve),
b) there must be at least 2 riders of Polish nationality and age under 21 years (junior riders)
c) there must be at least 4 riders of Polish nationality (junior riders not included)
d) there must be at least 1 rider of age under 24 years (U24 rider, nationality does not matter), this rider is not included in c) if he is Polish

The optimization models constructed belong to the class of linear integer programing models. Author propose two versions of the basic model, both based on RFR estimates for season 2020.

Version one. Integer programming model – maximizing the number of points scored by the team in the meeting (Model I):

$$TeamPts = \sum_{j=1}^{10} p_j x_j \rightarrow max \qquad (4.1)$$

subject to:

$$\sum_{j=1}^{10} v_j x_j \leq B \qquad (4.2)$$

$$\sum_{j=1}^{10} x_j \leq 8 \qquad (4.3)$$

$$x_5 + x_6 + x_7 \geq 4 \qquad (4.4)$$

$$x_3 + x_4 \geq 1 \qquad (4.5)$$

$$x_1 = 2 \qquad (4.6)$$

and

$$x_j \geq 0, x_j \in I. \qquad (4.7)$$

where: x_j- decision variable, number of riders of type *j* in a team; p_j- parameter, mean number of points delivered by rider of type *j*; c_j – parameter, cost of point delivered by rider of type *j*; B – constraint parameter, budget value allocated to cover payments for points, *j* = 1 … 10 – types of riders implied by Table 3.

Description of a model's structure is straightforward. 4.1 is an objective function with a value of a mean sum of points collected by a team. 4.2 is a budget constraint. 4.3 is a constraint representing point a) above, 4.4 representing c), 4.5 representing d) and equality 4.6 representing b). Boundary conditions 4.7 determine the membership of decision variables to a set of non-negative integers.

Second version of an optimization problem worth to consider is a decision problem where one seeks for minimal cost allocation in a presence of a constraint which determine a sport goal. It is reasonable to assume that the team look for at least 45 points in a meeting (this number of points guarantee a draw). With this assumption the version two o f a model takes a form:

Version two. Integer programming model – minimizing of cost of points (Model II)

$$Cost - \sum_{j=1}^{10} c_j x_j \rightarrow min \qquad (5.1)$$

Subject to:

$$\sum_{j=1}^{10} p_j x_j \geq 45 \qquad (5.2)$$

$$\sum_{j=1}^{10} x_j \leq 8 \qquad (5.3)$$

$$x_5 + x_6 + x_7 \geq 4 \qquad (5.4)$$

$$x_3 + x_4 \geq 1 \qquad (5.5)$$

$$x_1 = 2 \qquad (5.6)$$

and:

$$x_j \geq 0, x_j \in C. \qquad (5.7)$$

Definitions of variables and parameters of Model II are the same as in Model I. The structure of Model II differs from Model I in two elements. 5.1 is an objective function with a value of an average cost of points scored by the team in a meeting. Constraint 4.2 is replaced by the constraint 5.2, which guarantees at least 45 points collected by a team.

To validate the proposed Models author found optimal solutions of them. As a values of parameters p_j predictions from Table 3 were used. Due to the lack of exact knowledge of financial data, some arbitrary, albeit justified, assumptions were made regarding the c_j parameters' values. The values are normalized to range [0, 1] and reassembles possible "valuations of the rider" depending on nationality and age. The values *B* of the right-hand side of the budget constraint are set to simulate an actual or no budget constraint.

Optimal solution have been found using branch and bound method (Land, Doig, 1960) implemented in the Python programming language. The results of an optimizations contains Table 4.

Table 4. Optimization results

Variable name based on Table 3	Variable in the Models	p_j	c_j	optimal solution of Model I (budget constraint not binding)	optimal solution of Model I (budget constraint binding)	optimal solution of Model II
P21	x_1	2.232	0.4	2	2	2
I21	x_2	2.281	0.5	0	1	0
P24	x_3	5.132	0.65	0	0	0
I24	x_4	5.746	0.7	1	1	1
P40	x_5	6.710	0.8	0	0	0
P_over41	x_6	6.710	0.75	0	4	2
P32	x_7	7.226	0.85	4	0	3
I40	x_8	7.324	0.95	0	0	0
I_over41	x_9	7.324	0.9	0	0	0
I32	x_{10}	7.817	1	1	0	0
optimal value of the objective function (Points / Cost)				**46.92**	**39.33**	**5.55**
value of right hand side of a budget/point constraint				6	5	45
value of left hand side of a budget/point constraint for optimal solution				**5.9**	**5**	**45.30**

Discussion

Proper personnel and financial decisions (Pawłowski, 2020) play a key role in managing a club and a team in the team sports. In the conditions of progressing globalization of sport (Gulak-Lipka, 2020), such decisions become extremely complicated decisions. In speedway, team building decisions must take into account a number of factors, such as the expected point goal for the team, financial constraints, restrictions resulting from league regulations, and restrictions related to the availability of players of a certain nationality or age.

The paper develops novel method of analysis of historical impact of nationality/age structure of a team on efficiency and propose decision mechanism that could support managers in a team building process. Using the example of the highest level Polish speedway league, the author shows how important the individual predictors related to the nationality and age of the player were for the forecast average number of match points. It is clear now that the sets of predictors with the strongest impact on the score changed over the analysed sample. However, the influence of predictor *16 – 22* (regardless of nationality) was invariably high and negative over a whole sample. It means, that proper policy regarding junior riders could be very important factor in a success of a team.

The RFR estimation allows not only for importances' valuation but for prediction of mean number of points for some interpretable types of players, either. It is novel application of RFR in a sport domain. On a basis of such calculation (example contain Table 2 or Table 3) manager can assess approximate, mean "point value" of meaningful combinations of features. That features combinations are called as "types of a rider". For example, from Table 3 one can learn that valid feature combinations {Pol; 33-40} and {Pol; over 41} creates two types of rider which are equally efficient in terms of a mean sum of points collected.

RFR estimation provided a motive for the second stage of the study, which is the construction of the optimization model supporting the decision on the structure of the team registered for the meeting. Author proposes two version of optimization mechanism, both constructed in concordance with the championship of PGE Ekstraliga Regulation, valid for a season 2021. To validate the Models, optimal decisions have been found, based on numerical values of parameters partly taken from RFR estimation. Analysing optimal decision from Model I (Table 4) one can state that the optimal team consists of two junior riders, one foreign U24 rider, four Polish riders in the age between 25-32 years and one foreign rider in the age between 25-32 years. This team could collect 46.92 point per meeting on average. With assumed costs of riders, start of the teams in that composition cost 5.9 monetary units. When one assume that the club has more strict financial limits, and restrict budget value to 5 monetary units there is optimal decision either but that decision do not guarantee a success in a meeting (on average), as number of points collected amounts only 39.33. It leads to version two of optimization model – Model II. In this model, a minimum budget value is sought that will allow to achieve the sports goal, i.e. to collect the number of points that will allow at least a draw (it is 45 points). Solving Model II with the same parameters' values as in Model I one can obtain different optimal structure of the team and can learn the value of a budget which is necessary for maintaining that team (Table 4, last column).

Conclusions

Managing a sport team is an extremely difficult task due to many limitations imposed by the internal and external environment. One of the tasks faced by managers is to build the optimal composition of the team, taking into account the measures that define the sports goal. The article attempts to analyse the optimization process of building of a speedway team in the conditions of financial and organizational constraints.

The paper develops novel method of analysis of impact of various features connected with the rider on his efficiency and propose decision-making stack that could support managers in a team building process. The first component of the stack, the random forest algorithm, is for data mining and evaluation of importance of the individual predictors for the forecast of average number of match points. The use of the global and local interpretability of the random forest allows to reveal the determinants that allow the riders and the team to achieve a specific sports result. On a basis of the predictions generated by the algorithm manager can assess approximate, mean "point value" of a meaningful combinations of features that describe the rider. That features' combinations are called as "types of a rider". Finally, this step passes the parameters to next element of the stack which is an optimization model supporting the decision on the structure of the team registered for the meeting.

Optimization model (proposed in two variants) could help determine demographic and nationality structure of the team in concordance with regulation's and financial constraints.

It is shown that machine learning algorithms can be treat as a flexible tool not only for usual prediction task in sport but either as a source of information on historical behaviour of a sport teams, passing estimates to optimization mechanisms.

The proposed analytical framework - the Random Forest Regression model is a scalable tool in the sense that the set of predictors used in the study (national and demographic predictors) can be extended if appropriate statistical data are available. In that context, introducing telemetry service in Polish Ekstraliga opens new possibilities for a data driven methods.

The optimization models can also be developed by applying additional constraints, and above all by introducing parameters' values that are only known to the management of clubs (e.g. they are a trade secret). The Models I and II could be a base for simulation's tasks, either. It is hoped that the results of this study will serve as a decision-making tool for managers not only of speedway sports but other team disciplines.

References

Aćimović, D.; Špirtović, O.; Projević, A. (2013) The term of sport management and its significancein the context of application in the modern sport. Activities in Physical Education and Sport, vol. 3, no. 2, p. 238

Bejger, S., Elster, S. (2020). AI in economic decision making – how to assure a trust, *Economics and Law*, (19), 34, pp. 411 - 434.

Biau1, G., Scornet, E. (2016). A random forest guided tour, *TEST* 25:197–227, Springer

Breiman, L. (2001). Random Forests, *Machine learning*, 45(1), pp.5–32

Bunkera,R., Susnjak, T. (2019) The Application of Machine Learning Techniques for Predicting Results in Team Sport: A Review, arXiv:1912.11762v1

DeSensi, J.T.; Kelley, D.R.; Blanton, M.D.; Beitel, P.A. (1990) Sport management curricular evaluation and needs assessment. A multifaceted approach. Journal of Sport Management, 4, 31-58

Gulak-Lipka, P. (2020). Internationalization and managing diversity on the basis of professional basketball clubs, *Journal of Physical Education and Sport*, Vol.20 (6), Art 484, pp. 3591 - 3598

Hillier, F.S., Lieberman, G. J. (2001). *Introduction to operations research*, McGraw-Hill, New York.

Ho, T. K. (1998).The random subspace method for constructing decision forests, *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 20(8), pp.832–844

Fujii, K. (2021) Data-driven Analysis for Understanding Team Sports Behaviors, arXiv:2102.07545v2

Koerber, W. (2020), electronic document available at: https://po-bandzie.com.pl/zawodnicy-zrobili-z-nas-placzacych-egoistow-oto-nasze-koszty/

L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. (1984). *Classification and regression trees*, Belmont, CA: Wadsworth International Group

Land, H., Doig, A. G (1960). An automatic method of solving discrete programming problems, *Econometrica*, 28 (3). pp. 497–520

Lis, A., Tomanek, M. (2020). Sport management: Thematic mapping of the research field. *Journal of Physical Education and Sport*.Vol 20 (Supplement issue 2), Art 167, pp. 1201 – 1208

May, C. (1978). *Ride it ! The Complete Book of Speedway*, Haynes Publishers, UK.

Pawłowski, J. (2020) Financial condition of football clubs in the Polish Ekstraklasa, *Journal of Physical Education and Sport*, Vol 20 (Supplement issue 5), Art 385, pp 2839 – 2844

PGE Ekstraliga Statistics. (2021). retrieved from: https://speedwayekstraliga.pl/statystyki/frekwencja/?y=2019, on 15.02.2021

Quinlan, J. R. (1993). *C4.5: programs for machine learning*, volume 1. Morgan Kaufmann Publishers, Inc.

Polish team championship of PGE Ekstraliga Regulation. (2021). electronic document available at: https://www.pzm.pl/pliki/zg/zuzel/2021/Regulaminy/dmp2021_21.10.2020.pdf

Ribeiro, M. T., Singh, S., Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016