

Results of the FlyBall Race Analysis

Made by Dr. Sándor Csizmazia

sandor.csizmazia@t-online.hu

12.05.2024

A. Preparatory steps:

The received excel files: 00.EFC_versenyadatok.xlsx, 00.FLYBALLCZ_versenyadatok.xlsx, 00.Wild_Runners_Flyball_Team_kutyak_alapadatai.xlsx

had to be converted to distinguish between the red and blue coloured pitches in excel.

Result of the conversion: 00.EFC_versenyadatokM.xlsm, 00.FLYBALLCZ_versenyadatokM.xlsm

During processing, the incoming data had to be transformed to include a dog-related event in one row.

File: 01.aFlyBall.DataE.ipynb

The data cleaning and preparation can be found in the following files: 02.a.adattisztitas.ipynb, 03.adatelokeszites.ipynb

B. Analyses I.

Evaluation criteria in order of importance:

The following indicators are of particular interest for the annual evaluation of the Flyball competition:

1. Number of tournaments won by teams:

Indicates how many tournaments a team won during the year.

WildRunners	Maximini	17
WildRunners	Zephyr	13
WildRunners	hopefuls	13
WildRunners	dreamers	10
WildRunners	Nightfall	8
WildRunners	XXL	7
WildRunners	Wonders	2

Maximini excelled in several divisions and showed a balanced performance throughout the year. Zephyr and Nightfall stand out in a similar way.

Their Min and average total time contributed significantly to their podium finish.

The XXL, the Wonders ran very poor average times, had very poor max totals, competed in few races so they scored significantly less points and won races.

2. The number of points scored by the teams:

The total number of points that are calculated based on race finishes.

WildRunners	Zephyr	51.0
WildRunners	Maximini	47.0
WildRunners	hopefuls	46.0
WildRunners	Nightfall	37.0
WildRunners	dreamers	36.0
WildRunners	XXL	19.0
WildRunners	Wonders	3.0

3. Number of points scored by each dog in a competition, their progress

This is important because it allows you to compare the performance of individual dogs.

The analysis contains information on this, here I will only present the outperformances in the next section.

Here I will refer to the series of graphs entitled Individual development of dogs, which shows that:

- a. Bella, Bree, Brooke, Bailey, Hank, Fleur, Hector, Maggie, Tyna : they competed in a balanced way, some of them improved during the year

Pay special attention to their starters, because they also play a significant role in the dogs' excellent performance!

- b. Alex, Affie, Ashley, Bonnie, Dexter, Dixie, Enola, Falco, Gerry, Gordon, Jasper, Kesha, Mason, Moon, Nina, Sara, Sunny, Ted:

Started well but declined or didn't compete from mid-year.

For them, better results can definitely be achieved with training, medical assistance and training of starters, so they need to be addressed, invested in.

- c. Abby, Agnes, Brix, Easy, Henry, Jessie, Kent, Milo, Punk, Rex, Riley

Here we need to consider not only the dogs but also the role of the starters, as this is closely linked to the dog's performance.

- d. There are dogs that have performed well after a significant absence, probably due to injury, illness, other reasons, but have performed well on return, Angie, Brix, Bruno, Cookies, Dusty, Easy, Henry, Rocky, Scotty, Sirius, Speedy, Them

4. The sum of the points scored by each dog during the year:

This shows how many points a dog has scored in total during the year.

This is the number of dogs in a dog's year:

Hector	8.0
Alfie	8.0
Henry	6.0
Them	5.0
Riley	5.0
Negro	5.0
Moon	5.0
Maggie	5.0
Enola	5.0

weak:

Punk	2.0
Kent	2.0
Easy	2.0
Rex	1.0
Milo	1.0

important aspects are:

- average race time
- stability of the team composition
- the error rate

These are shown in the analysis, but I will not go into them.

File: 04.elemzes.alapinfok.ipynb

C. Analyses II.

During the competition, we were asked questions which we had to answer in turn.

These were:

1. would like to know if the dogs' running performance is significantly affected by the colour of the track they run on

The task can also be seen as a classification task. If classes can be found, then the colour of the track has an effect on performance.

Result: the start time, individual run time, group time, group winning chance of the dogs running on the track are independent of the colour of the track.

There is a weak correlation: start times of red track runners are worse (could be a coincidence)

File: 04.elemzes.palya.szine.ipynb

2. would also be curious to know if there is a significant effect on the dogs' running performance of switching to a male/ratless mare/female

The task could be seen as a classification task. If classes can be found, then there is an effect of the previous runner on the performance of the subsequent one.

Result: independent of the previous runner, the individual running time, the group time, the group winning chance of the dogs running on the course.

For start times, the start times for male and female dogs are worse than the rest.

There is a weak correlation: the running times (dog) of dogs following a gelding are sometimes worse.

File: 04.elemzes.ivaros.ipynb

3. or he would like to see all the previous analyses (the previous basic analyses) linked to competitions (and venues and dates), because that would tell him more about the information

Result: the previous analyses broken down by run, place, when variables, the findings are similar to the Round II:

Bad start time usually means a loss (so this and therefore the role of the starter seems to be a priority in the race) - regardless of race, time, date

MIN run time (total) does not necessarily mean a win - regardless of race, time, date

MAX running time (total) usually means a loss - regardless of race, time, date

File: 04.elemzes.alapinfok.ipynb

D. Analyses III.

The weather data were retrieved with this file: 03.adatelokeszites.ipynb

The effects of the following weather conditions were investigated:

1. wind speed:

(Winners: 2, Losers: 0, Ties: 1)

Wind strength increases with class code. generally teams perform better in light winds (<8.21),

some teams perform worse in strong winds: 'WildRunners Wonders'

File: 04.elemzes.szelerosseg.ipynb

2. Wind direction

(Winners: 2, Losers: 0, Ties: 1)

Wind direction: {12: 'W', 13: 'WNW', 6: 'NW', 2: 'N', 5: 'NNW', 3: 'NE', 4: 'NNE', 14: 'WSW', 11: 'SW', 8: 'SE', 1: 'ESE', 0: 'E', 9: 'SSE', 10: 'SSW', 7: 'S'}

There is a close relationship between wind direction and team victory, with a nice separation between the wind directions of winners and losers

but there are teams that run well or badly regardless of wind direction

File: 04.elemzes.szelerany.ipynb

3. temperature:

Winners: 2, Losers: 0, Ties: 1

The temperature increases with the class code. There is a weak correlation between dogs' performance and temperature

file: 04.elemzes.homerseklet.ipynb

4. humidity:

Winners: 2, Losers: 0, Ties: 1

The humidity increases with the class code. There is very little or no correlation between the dogs' performance and humidity

file: 04.elemzes.paratartalom.ipynb

5. precipitation

Winners: 2, Losers: 0, Ties: 1

The chance of precipitation increases with the class code. There is very little or no correlation between the dogs' performance and the chance of precipitation

file: 04.elemzés.csapadek.ipynb

E. Feature Importance

Using the KN Classifier, I found the following ordering of feature importance:

1) W_max	0.271168
2) L_count	0.103817
3) when	0.094101
4) date	0.083657
5) division	0.066780
6) who	0.061452
7) hurdless	0.042964
8) place	0.034424
9) total	0.031380
10) temp_c	0.025201

I did the same measurement with the H2O analyzer that I learned in our Dataclub, and it showed roughly similar results.

File: 04.elemzes.feature.importance.ipynb

F. Definition of new indicators

I defined two important new indicator families:

a. W_count, L_count, T_count: sum of the results of the teams in 'datumlabel', 'wholabel', 'division'

b. W_max: the largest of the W_counts in 'datumlabel', 'division' (winner of the race on the given day and in the given division)

File: 03.adatelokeszites.ipynb

G. Creating a forecast model

Our data can also be treated as time series data.

Using the defined indicators and the feature importance selected using the Rolling Window technique familiar from Dataclub

to create a model that can be used to provide forecasts for the coming periods.

Using the facilities of H2o, I defined the model and saved it.

File: 05.H2O.Forecast.ipynb

./h2o_best.05.forecast.V01/GBM_grid_1_AutoML_3_20240512_163515_model_22

Thanks to the organizers for the interesting exercises! We learned a lot from them.

H. Ongoing Works

Forecast and modelling continues (outside the competition).

Bibliography:

1. Raschka, Sebastian, and Vahid Mirjalili. *Python Machine Learning, 3rd Ed.* Packt Publishing, 2019.
2. Meenakshi Gupta and Latika Singh, *Horse Race Results Prediction Using Machine Learning Algorithms With Feature Selection*,
International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING,
Volume: 11 Issue: 9, 2023