

A FlyBall Verseny eredményei

Készítette: Csizmazia Sándor

sandor.csizmazia@t-online.hu

2024.05.12

A. Előkészítő lépések:

A kapott excel file-okat: 00.EFC_versenyadatok.xlsx, 00.FLYBALLCZ_versenyadatok.xlsx, 00.Wild_Runners_Flyball_Team_kutyak_alapadatai.xlsx

át kellett alakítani, hogy az excelben a piros és kék színnel jelzett pályák megkülönböztethetők legyenek.

Az átalakítás eredménye: 00.EFC_versenyadatokM.xlsm, 00.FLYBALLCZ_versenyadatokM.xlsm

A feldolgozás során úgy kellett átalakítani a bejövő adatokat, hogy egy sorában egy kutyával kapcsolatos esemény szerepeljen.

File: 01.aFlyBall.DataE.ipynb

Az adattisztítás és előkészítés a következő file-okban található: 02.a.adattisztitas.ipynb, 03.adatelokeszites.ipynb

B. Elemzések I.

Az értékelési szempontok fontossági sorrendben:

A Flyball verseny éves értékelésénél az alábbi mutatók mindenképp érdekesek:

1. A csapatok által megnyert versenyek száma:

Megmutatja, hogy egy csapat mennyi versenyt nyert meg az év során.

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

A Maximini több divisio-ban is kiválóan szerepelt és az egész év folyamán kiegyensúlyozott teljesítményt mutatott. Hasonlóan kiemelkedik a Zephyr és a Nightfall csapat.

A Min és átlagos total idejük jelentősen hozzásegítette őket a dobogós helyhez.

Az XXL, a Wonders nagyon rossz átlagidőket futottak, nagyon rossz max totalokat szereztek, keves versenyen indultak így jelentősen kevesebb pontot és verseny nyerést is értek el.

2. A csapatok által szerzett pontok száma:

Az összpontszám, amelyet a versenyeken elért helyezések alapján számítanak ki.

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. Az egyes kutyák által egy versenyen szerzett pontok száma, fejlődésük

Ez azért fontos, mert lehetővé teszi az egyes kutyák teljesítményének összehasonlítását.

Az elemzés tartalmaz erre vonatkozó információt, itt csak a következő pontban a kiugró éves teljesítményeket mutatom be.

Itt a kutyák egyéni fejlődése c. ábraszorozatot említtem meg, ami azt mutatja, hogy:

- a. Bella, Bree, Brooke, Bailey, Hank, Fleur, Hector, Maggie, Tyna : kiegyensúlyozottan versenyeztek, volt aki fejlődött az év során
Kiemelt figyelmet fordítsunk az indítóikra, mert nekik is jelentős szerepük van a kutyák kiváló teljesítményében!
- b. Alex, Affie, Ashley, Bonnie, Dexter, Dixie, Enola, Falco, Gerry, Gordon, Jasper, Kesha, Mason, Moon, Nina, Sara, Sunny, Ted:
jól indultak, de év közepétől csökkent a teljesítményük vagy nem is versenyeztek.
Náluk edzéssel, orvosi segítséggel, az indítók képzésével mindenképp jobb eredményeket lehet elérni,
így foglalkozni kell velük, be kell ruházni az esetükben.
- c. A következő kutyák esetén el kell gondolkodni, hogy megváltjunk-e tőlük: Abby, Agnes, Brix, Easy, Henry, Jessie, Kent, Milo, Punk, Rex, Riley
Itt nem csak a kutyák, hanem az indítók szerepét is meg kell vizsgálni, mert ez szoros kapcsolatban van a kutya teljesítményével.
- d. Vannak kutyák, akik jelentős hiányzás után is jól teljesítettek, valószínű, hogy sérülés, betegség, egyéb okok miatt nem szerepeltek, de visszatéréskor jól teljesítettek, meg kell vizsgálni a helyzetüket: Angie, Brix, Bruno, Cookies, Dusty, Easy, Henry, Rocky, Scotty, Sirius, Speedy, Them

4. Az egyes kutyák által az év során szerzett pontok összege:

Ez mutatja, hogy egy kutya mennyi pontot ért el összesen az év során.

kemelkedő:

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

gyenge:

Punk	2.0
Kent	2.0
Easy	2.0
Rex	1.0
Milo	1.0

fontos szempontok még:

- átlagos versenyzési idő
- a csapat összetételének stabilitása
- a hibaarány

Ezekre találunk diagramokat az elemzésben, de nem térek ki rá.

File: 04.elemzes.alapinfok.ipynb

C. Elemzések II.

A verseny során kérdéseket kaptunk, amiket sorra meg kellett válaszolnunk.

Ezek a következők:

1. szeretné tudni, hogy a kutyák futási teljesítményére szignifikáns hatással van-e az, hogy melyik színű pályán futnak

A feladat egy klasszifikációs feladatként is felfogható. Ha sikerül osztályokat találni, akkor hatással van a pálya színe a teljesítményekre.

Eredmény: független a pálya színétől a pályán futó kutyák start ideje, egyéni futási ideje, csoport ideje, csoport nyelési esélye.

Gyenge összefüggés van: a piros pályán indulók start idői rosszabbak. (véletlen is lehet)

File: 04.elemzes.palya.szine.ipynb

2. kíváncsi lenne arra is, hogy a kutyák futási teljesítményére szignifikáns hatással van-e az, hogy ivaros/ivartalan kanra/szukára vált rá a kutya

A feladat egy klasszifikációs feladatként is felfogható. Ha sikerül osztályokat találni, akkor hatással van az előző futó a rákövetkező teljesítményére.

Eredmény: független az előző futótól a pályán futó kutyák egyéni futási ideje, csoport ideje, csoport nyeresi esélye.

A start időknél az ivaros kanra és szukára következők start idői rosszabbak a többihez képest.

Gyenge összefüggés van: az ivartalan kutyákra következők futás idői(dog) néha rosszabbak.

File: 04.elemzes.ivaros.ipynb

3. illetve szeretné az összes eddigi elemzést (a korábbi alap elemzéseket) versenyekhez (és helyszínekhez, és dátumokhoz) rendelve látni, mert úgy többet mondanának neki az információk

Eredmény: a korábbi elemzéseket run, place, when változókra részletezve, a megállapítások hasonlóak a II. fordulóbeliekhez:

A rossz start idő általában veszteséget jelent (így ez és emiatt az indító szerepe kiemeltnek tűnik a versenyben) - versenytől, helyszíntől, dátumtól függetlenül

A MIN futási idő (total) nem feltétlen jelent győzelmet - versenytől, helyszíntől, dátumtól függetlenül

A MAX futási idő (total) többnyire veszteséget jelent - versenytől, helyszíntől, dátumtól függetlenül

File: 04.elemzes.alapinfok.ipynb

D. Elemzések III.

Az időjárás adatok lekérése ezzel a file-al történt: 03.adatelokeszites.ipynb

A következő időjárás körülmányek hatásait vizsgáltam meg:

1. szélerősség:

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

Szél erősség nő az osztály kódjával. általában gyenge szélben (<8,21) jobban teljesítenek a csapatok,

van olyan csapat aki erős szélben rosszabbul teljesít: 'WildRunners Wonders'

file: 04.elemzes.szelerosseg.ipynb

2. szélirány:

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

Szélirány: {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'}

A szélirány és a csapat győzeleme között szoros kapcsolat van, szépen

elkülönülnek a győztesek és a vesztesek szélirányai
ám van olyan csapat ameyik széliránytól függetlenül fut jól vagy rosszul
file: 04.elemzes.szelirany.ipynb

3. hőmérséklet:

Winners: 2, Losers: 0, Ties: 1
A hőmérséklet nő az osztály kódjával. Gyenge összefüggés van a kutyák teljesítése és a hőmérséklet között
file: 04.elemzes.homerseklet.ipynb

4. páratartalom:

Winners: 2, Losers: 0, Ties: 1
A páratartalom nő az osztály kódjával. Nagyon gyenge összefüggés van vagy nincs is a kutyák teljesítése és a páratartalom között
file:04.elemzes.paratartalom.ipynb

5. csapadékmennyiség

Winners: 2, Losers: 0, Ties: 1
A csapadék esélye nő az osztály kódjával. Nagyon gyenge összefüggés van vagy nincs is a kutyák teljesítése és a csapadék esélye között
file: 04.elemzés.csapadek.ipynb

E. Feature Importance

A KN Classifier segítségével a következő sorrendet találtam a jellemzők hatásfokára:

1) W_max	0.271168
2) L_count	0.103817
3) when	0.094101
4) datum	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

A Dataklubban megismert H2O elemzővel is elvégeztem a mérést, s nagyjából hasonló eredményt mutatott az is.

file: 04.elemzes.feature.importance.ipynb

F. Új mutatók definiálása

két fontos új mutatócsaládot definiáltam:

- W_count, L_count, T_count: a csapatok eredményeinek összege
'datumlabel', 'wholabel', 'division' bontásban
- W_max: a W_count-k közül a legnagyobb 'datumlabel', 'division' bontásban (a verseny nyertese az adott napon és divízióban)

File: 03.adatelokeszites.ipynb

G. Forcast model létrehozása

Adataink kezelhetők idősoros adatokként is.

A definiált mutatók és a feature importance kiválasztott jellemzők segítségével a Dataklubban megismert Rolling Window technika segítségével

modellt tudunk készíteni, amellyel előrejelzéseket tudunk adni az elkövetkező időszakokra.

A H2o lehetőségeit használva defináltam a modellt és mentettem el.

file: 05.H2O.Forecast.ipynb

./h2o_best.05.forecast.V01/GBM_grid_1_AutoML_3_20240512_163515_model_22

Köszönet a szervezőknek az érdekes feladatokért! Sokat tanultunk belőlük.

A Forcast és modellkészítés tovább folyik (versenyen kívül).