

```

---
output:
  pdf_document: default
  html_document: default
---

```

```
# Zárthelyi dolgozat
```

```
2022. április 22. Péntek
```

- Az érdemjegy az összes feladat elvégzésének eredményeképpen kerül meghatározásra.
- Maximális pontszám 100 pont.
- A vizsga megoldására 90 perc áll rendelkezésre.
- A félévközi aláírás megszerzéséhez legalább a 2-es (elégséges) szint elérése szükséges, ezen felül a kapott pontszám nem befolyásolja sem az aláírás megszerzését, sem a vizsgajegy meghatározását.
- A legutolsó mentés időpontja nem lehet későbbi, mint a dolgozat megkezdése +90 perc!!!
- Együttműködés gyanúja esetében az elért eredményt szóban fogjuk ellenőrizni!

```

91 -      5 (jeles)
81 - 90   4 (jó)
61 - 80   3 (közepes)
41 - 60   2 (elégséges)
0  - 40   1 (elégtelen)

```

```
## 1. R alapok (15 pont)
```

```
### a) Hozzon létre egy 3x4 méretű mátrixot 1 és 9 között egyenletes eloszlás szerint generált számokkal, majd az apply függvény segítségével soronként határozza meg az elemek szorzatát! (5 pont)
```

```

```{r}
```

```

```
### b) Az 1,2,3,4...9 számsorból válasszon ki 5 értéket úgy, hogy az értékek ismétlődhessenek! Ezek között milyen egyedi értékek szerepelhetnek? (5 pont)
```

```

```{r}
```

```

```
### c) Hozzon létre egy 8 elemű faktorizált vektort "a", "b" és "c" értékekkel! A faktor milyen indexeket tartalmaz? (5 pont)
```

```

```{r}
```

```

```
## 2. Vizualizáció (25 pont)
```

```
### a) A "ggplot2" könyvtárból töltsse be az "mpg" adathalmazt egy munka változóba! (2 pont)
```

```

```{r}
```

```

```
### b) Ábrázolja a négyhengeres autók városi fogyasztásának hisztogramját, valamint sűrűségfüggvényét egy ábrában! Az ábrát és tengelyeket feliratozza, valamint színezza a hisztogramot és a sűrűségfüggvényt! (10 pont)
```

```

```{r}

```

```
...
```

```
c) Egy ábrán ábrázoljon két ponthalmazt! A hengerúrtartalom függvényében a városi és az országúti fogyasztásokat! (13 pont)
```

- Az egyik ponthalmazt zöld pontokkal, a másik ábrát kék keresztekkel jelenítse meg!
- Figyeljen az egymást fedő pontokra!
- Az ábrát lássa el feliratokkal!
- Használjon jelmagyarázatot!

```
```\r}
```

```
...
```

```
## 3. Adatok megismerése (15 pont)
```

```
Végezze el az iris adathalmaz "setosa" fajtája "Sepal.Length" értékeinek normalitás vizsgálatát!
```

```
### a) Töltse be az iris adathalmazt és szűrje az adatokat a "setosa" fajtára és a "Sepal.Length" értékekre. (3 pont)
```

```
```\r}
```

```
...
```

```
b) Végezzen vizuális normalitás vizsgálatot sűrűség függvény segítségével! Az eredményt pár szóban értékelje! (6 pont)
```

```
```\r}
```

```
...
```

```
### c) Végezze el az adatok normalitás vizsgálatát tetszőleges statisztikai paraméterek felhasználásával! A kapott eredményeket értékelje pár szóban! (6 pont)
```

```
```\r}
```

```
...
```

```
4. Regresszió - Predikció (20 pont)
```

```
Végezzen lineáris regressziót az iris adathalmaz adott két paramétere között!
```

```
a) Hozzon létre két adatvektort, mely az iris adathalmaz "setosa" fajták "Sepal.Length" és "Petal.Length" adatait tartalmazza. (2 pont)
```

```
```\r}
```

```
...
```

```
### b) A két adatvektorra hozzon létre lineáris regressziós modellt, ahol a "Sepal.Length" értéke a deskriptor! (6 pont)
```

```
```\r}
```

```
...
```

```
c) Egy ábrán ábrázolja az eredeti ponthalmazt, valamint a lineáris regresszió eredményét! (6 pont)
```

```
```\r}
```

```
...
```

d) Mekkora lenne a `Petal.Length` értéke, ha a `Sepal.Length` 6-os értéke mellett? (6 pont)

```
```{r}
```

```
```
```

5. Osztályozás (25 pont)

Az osztályozási feladat a `scat` adathalmazon (megtalálható a `caret` csomagban) történik. Az adathalmaz három állatfajta (bobcat,coyote,gray_fox) székletadatait tartalmazza.

a) Töltsük be az adathalmazt!

```
```{r}
```

```
```
```

b) Készítsünk egy adathalmazt, amely csak a `Species,Site,Location,Length,Diameter,Mass` változókat tartalmazza! Amennyiben az adathalmaz tartalmaz `NA` értékeket, távolítsuk el azokat a sorokat, amelyekben megtalálhatók! (5 pont)

```
```{r}
```

```
```
```

c) Válasszuk szét az adathalmazt teszt és tanuló adathalmazra! A tanuló adathalmaz az adatok 70%-a legyen! Figyeljünk oda, hogy osztályarányosan végezzük a szétválasztást a `Species` faktor alapján! (5 pont)

```
```{r}
```

```
set.seed(0)
```

```
```
```

d) Készítsünk egy döntési fa osztályozást az adathalmazra! Ehhez használjuk célváltozónak a fajt (`Species`), és az összes többi változót jellemzőnek! Keressük meg a fa optimális mélységét bootstrap validációval 1 és 5 mélység között! Melyik lett az optimális mélység? (4 pont)

```
```{r}
```

```
set.seed(0)
```

```
```
```

e) Ábrázolja a döntési fát! Melyik változó volt a legfontosabb az osztályozás során? (3 pont)

```
```{r}
```

```
```
```

f) Készítsük el az osztályozáshoz tartozó keveredési mátrixot, mind a tanító, mind a teszt adathalmazhoz! Értékelje ki pár mondatban az osztályozás eredményét! (4 pont)

```
```{r}
```

```
```
```

g) Javítható az osztályozás, ha a tanító adathalmazt kiegyensúlyozzuk? Mutassa meg, hogy igen vagy sem (a minőség értékeléséhez használjuk a pontosság (accuracy) mutatót)! (4 pont)

```
```{r}
```

```
```
```