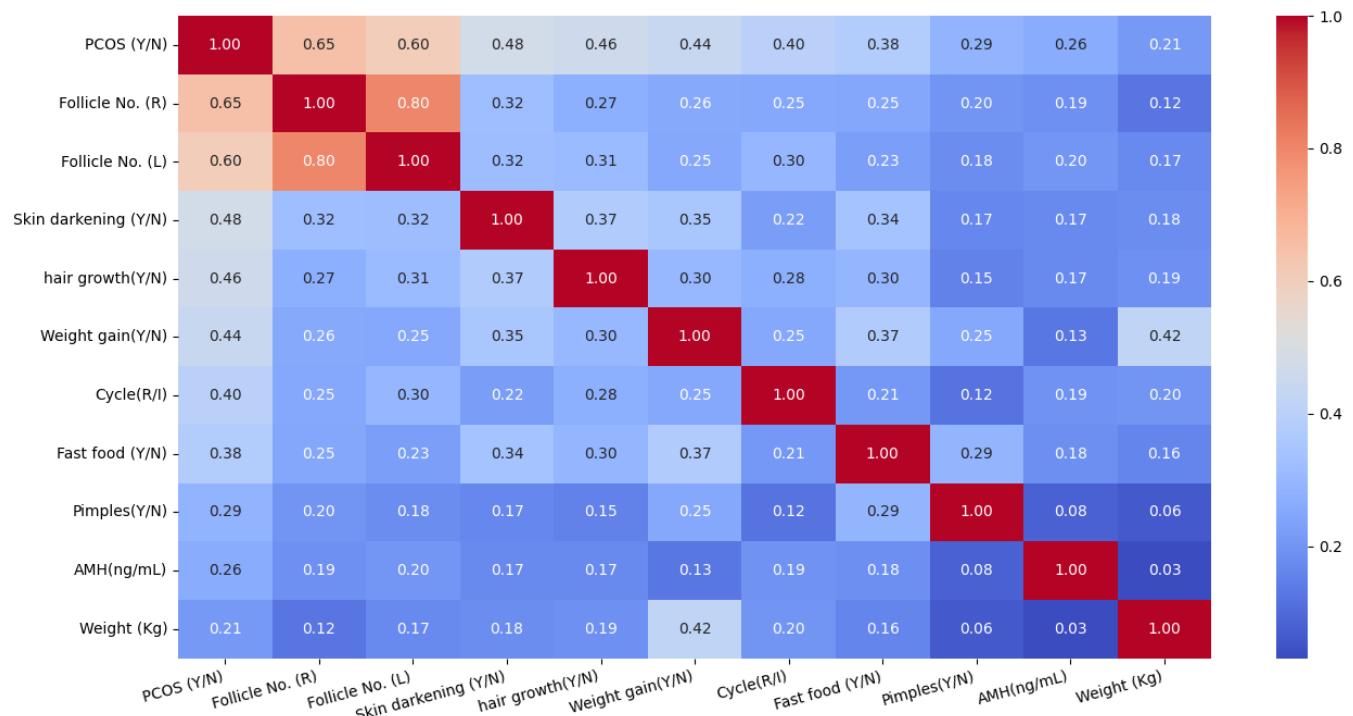


Explorarea legaturilor dintre o serie de factori medicali, de stil de viata si PCOS

Pentru analiza descriptiva a unui grafic am ales un set de date de pe Kaggle care contine informatii despre paciente care sufera de Polycystic ovary syndrome (PCOS). Sindromul ovarelor polichistice este o afectiune des intalnita in randul femeilor cauzand aparitia mai multor foliculi sau chisturi mici in ovare, cat si a dezechilibrelor hormonale. De asemenea, PCOS este asociat si cu alte conditii mai severe, precum anxietate, depresie, diabet de tip 2, rezistenta la insulina sau boli cardiovasculare.

Setul de date contine un fisier Excel in care sunt introduse date fizice despre paciente si mai multi parametri clinici. Pentru inceput am colectat si pre-procesat datele. Am citit datele din Excel utilizand biblioteca pandas in limbajul Python si am eliminat date mai putin importante, sau date care se calculeaza pe baza altora, deoarece exista deja o legatura directa intre anumite variabile si ne poate influenta analiza, scopul fiind de a gasi legaturi care nu sunt foarte evidente. Am completat datele care au valori lipsa (NaN) si am observat ca pentru variabila Fast Food exista o valoare NaN pe care am completat-o cu mode.



In figura de mai sus este reprezentat graficul heatmap a corelatiilor dintre PCOS si o serie de factori medicali si de stil de viata. In cadrul setului de date numarul de coloane este de 45, insa pentru a evidenția care ar fi factorii care au o legatura mai puternica cu PCOS am ales un prag de 0.2, ramanand doar 10 factori care au o legatura cat mai pozitiva cu diagnosticul de PCOS. O culoare mai calda indica o corelatie pozitiva mai puternica, in timp ce culorile reci de albastru indica o corelatie pozitiva dar mai slaba.

Intr-adevar, se confirma faptul ca un numar crescut de foliculi pe ovarul drept, cat si cel stang este asociat cu PCOS. Mai mult, se pare ca apare o corelatie destul de puternica intre foliculii de pe ovarul drept si cel stang, cel mai probabil cineva care are chisturi pe oricare dintre ovare poate sa aiba si pe celalalt.

Urmatoarea variabila care are o corelatie semnificativa cu PCOS este skin darkening, care este adesea si un simptom al rezistentei la insulina, indicand legatura dintre PCOS si afectiuni ale metabolismului.

De asemenea, se pare ca si stilul de viata influenteaza PCOS, indicand faptul ca o alimentatie nesanatoasa, consumul de Fast Food, si greutatea corporala pot influenta sindromul ovarelor polichistice.

Heatmap-ul prezentat ne ofera o imagine de ansamblu asupra conexiunilor dintre diferitele aspecte ale acestei afectiuni, cat si o oportunitate de a explora datele si de a gasi potentiale asocieri intre variabile, insa corelatie dintre ele nu se poate stabili si cauzalitatea. Un simptom poate sa nu cauzeze neaparat un alt simptom.

Anexa

Sursa date

<https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos>

Cod

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

pd.set_option("display.max_columns",100)
PCOS_data =
pd.read_excel(r'C:\Users\Ruxi\Desktop\An1Sem1\SA\PCOS_data_without_infertility.xlsx',sheet_name="Full_new")
print(PCOS_data.columns)
print(PCOS_data.shape)

PCOS_data = PCOS_data.drop(["Sl. No","Patient File No.", "Unnamed: 44","FSH/LH","Waist:Hip Ratio","BMI","Marraige Status (Yrs)"],axis = 1)

mode_value = PCOS_data['Fast food (Y/N)'].mode().iloc[0]
PCOS_data['Fast food (Y/N)'].fillna(mode_value, inplace=True)
#verificare valori nenule
print(PCOS_data.isnull().sum())
#date numerice pentru corr
PCOS_data = PCOS_data.apply(pd.to_numeric, errors= 'coerce')
print(PCOS_data.head())

corr_matrix= PCOS_data.corr()
# print(corr_matrix)
prag = 0.2
filtre = np.abs(corr_matrix["PCOS (Y/N)"]) > prag
corr_features = corr_matrix.columns[filtre].tolist()
corr_features = corr_matrix.loc[corr_features, 'PCOS (Y/N)'].abs().sort_values(ascending=False).index.tolist()

sns.heatmap(PCOS_data[corr_features].corr(), cmap="coolwarm", annot=True, fmt=".2f")
plt.xticks(rotation=15,ha="right")
plt.show()
```