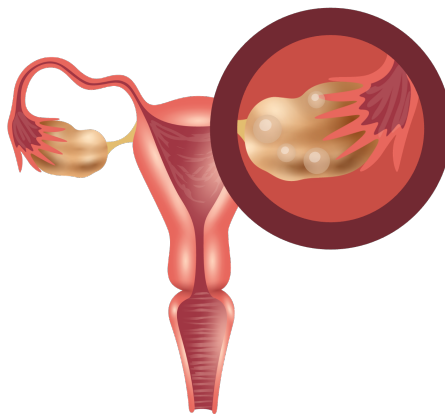


Departamento de Ciências Biomédicas da Universidade de Aveiro

Mestrado em Bioinformática Clínica

Programação e Algoritmos em Ciências

Síndrome do Ovário Policístico (PCOS)



Adriana Gomes | 117348 | adriana.gomes@ua.pt

Ana Rita Silva | 121764 | ar.m.silva@ua.pt

15 de Dezembro de 2023

1 Base de Dados

O presente projeto foi desenvolvido na linguagem Python 3 e tem como intuito a criação de um menu de utilização adaptado à base de dados escolhida. Para tal, requeria-se a leitura dos dados presentes no *dataset*, apresentá-los de forma pertinente, visual e, por fim, que permitisse guardar em ficheiro os resultados obtidos.

Com recurso à plataforma Kaggle foi selecionada uma base de dados *open-source* sobre Síndrome do Ovário Policístico (PCOS) intitulada **PCOS Dataset** disponibilizada por Shreyas Vedpathak. A base de dados é constituída 45 colunas em que 3 delas foram removidas – uma coluna não continha dados e as outras duas eram referentes à numeração ordenada dos indivíduos. Assim, a base de dados apresenta 42 variáveis e conta com 541 observações, sendo 177 pacientes diagnosticadas com a Síndrome do Ovário Policístico (PCOS). O conjunto das variáveis disponibilizadas reúne diversas características, hábitos e factores a ter em consideração na saúde da mulher.

Variável	Tipo	Caracterização
PCOS (Y/N)	Integer	Binomial. Presença (1) e Ausência da Síndrome (0)
Age (yrs)	Integer	Idade do indivíduo (em anos)
BMI	Float	Índice de Massa Corporal
Cycle length(days)	Integer	Duração da menstruação (em dias)
Pregnant(Y/N)	Integer	Binomial. Paciente grávida (1) ou não se encontra grávida (0)
No. of abortions	Integer	Número de abortos sofridos.
FSH(mIU/mL)	Float	Concentração da Hormona Folículo-Estimulante (mili unidades internacionais por mililitro)
LH(mIU/mL)	Float	Concentração Hormona Luteinizante (mili unidades internacionais por mililitro)
FSH/LH	Float	Razão entre as Hormonas Folículo-Estimulante e Luteinizante
TSH (mIU/L)	Object → Float	Concentração da Hormona Estimuladora da Tiróide (mili unidades internacionais por litro)
AMH(ng/mL)	Float	Concentração da Hormona Anti-Mulleriana (nanograma por mililitro)
Vit D3 (ng/mL)	Float	Concentração de Vitamina D (nanograma por mililitro)
PRG(ng/mL)	Float	Concentração de Progesterona (nanograma por mililitro)
RBS(mg/dl)	Float	Níveis de glucose no sangue (miligrama por decilitro)
Weight gain(Y/N)	Integer	Presença (1) e ausência (0) de ganho de peso.
hair growth(Y/N)	Integer	Presença (1) e ausência (0) de crescimento de pelos.
Reg.Exercise(Y/N)	Integer	Prática regular (1) ou não (0) de Exercício Físico
Follicle No. (L)	Integer	Número de Folículos presentes no ovário esquerdo.
Follicle No. (R)	Integer	Número de Folículos presentes no ovário direito.
Endometrium (mm)	Float	Espessura do endométrio (milímetros)

Figura 1: Resumo da Caracterização das Variáveis em Estudo.

No decorrer da elaboração do projeto utilizou-se 20 variáveis que se considerou mais pertinentes tendo em conta o enquadramento do trabalho. No quadro acima (Figura 1) encontra-se representado o nome das variáveis, assim como o respetivo tipo e caracterização.

A PCOS é uma síndrome de diagnóstico pouco linear e, muitas vezes, difícil de obter. Como tal, pretende-se apresentar ao utilizador uma oportunidade de ver os seus valores (previamente obtidos nas suas análises) interpretados, tendo em conta os valores de referência para cada parâmetro disponibilizado. Em adição, tem-se também como objetivo realizar uma análise exploratória dos dados para permitir ao utilizador - sob a forma de gráficos e tabelas- perceber qual a incidência de diversos parâmetros na avaliação da presença de PCOS. Por fim, é de salientar que o programa desenvolvido tem como alvo o público em geral.

2 Programa

O programa permite ao utilizador navegar por uma análise exploratória de dados relativos à doença PCOS, tendo por base a pesquisa de tabelas e gráficos que mostram relações entre diferentes parâmetros. Além disso, é composto por uma secção que proporciona a introdução de dados acerca de determinados parâmetros, sendo que é devolvido um relatório clínico com a avaliação tendo em conta intervalos de referência. Desta forma, o utilizador é capaz de, no final da sua avaliação clínica, perceber se os seus valores são considerados normais ou anormais.

Este programa é apresentado numa pasta zip, que deve ser extraída na sua totalidade para uma pasta disponível para o efeito (escolhida pelo utilizador). Esta pasta zip contém um ficheiro de texto (Packages a instalar.txt), que mostra quais os pacotes a instalar, de modo a que o programa funcione corretamente. Ademais, a pasta possui o programa em si (Programa.py), que apresenta um menu interativo que pode ser explorado pelo utilizador.

A partir do menu, o utilizador pode aceder: ao botão "**Parâmetros**", que lhe permitirá aceder a tabelas com médias, mínimos e máximos referentes a diversos parâmetros; ao botão "**Gráficos**", que mostra diversos tipos de gráficos avaliando a relação entre parâmetros relativos à doença; ao botão "**Tabelas**", que disponibiliza frequências relativas e absolutas sobre parâmetros implicados na PCOS; ao botão "**Avalie situação clínica**", onde é possível realizar a introdução de dados para que depois seja criado um relatório clínico, que pode ser consultado em formato PDF; e, por fim, um botão denominado "**Sair do Programa**", que termina o programa, fechando todas as janelas abertas pelo utilizador.

2.1 Menu

Este programa dispõe de um menu interativo criado através do pacote tkinter ("interface Tk"). Este pacote é a interface padrão do Python para o toolkit GUI Tcl/Tk. O pacote tkinter foi incorporado no programa usando os comandos "from tkinter import *" para o módulo principal e "import tkinter as tk" para o módulo tk.

Deste modo, foi utilizado para criar um widget Tk, isto é, a janela principal, que foi manipulada para exibir informações complementares em janelas pop-up. Para aceder a cada janela, foram criados botões aos quais foi adicionado um comando que chama uma função específica. Esses botões foram personalizados em texto, formato e cor.

Os comandos atribuídos aos botões são funções que abrem novas janelas interativas exibindo novos botões que permitem ao utilizador navegar por diferentes variáveis. Portanto, é nestas janelas pop-up que o utilizador consegue aceder a gráficos, tabelas e ao local de introdução de dados.

2.2 Gráficos

De modo a ser possível criar gráficos, começou-se por realizar a leitura da base de dados escolhida. De seguida, sucedeu-se o tratamento dos dados, onde foram eliminadas a primeira, segunda e última colunas. Além disso, foi necessário converter as variáveis AMH e beta HCG para numéricas. De modo a eliminar valores em falta, foi realizada uma procura desses valores e, de seguida, foram preenchidos pela média da coluna onde se encontravam.

Terminado o tratamento dos dados, passou-se à execução dos gráficos. Neste caso, foram necessários os pacotes matplotlib, seaborn e pandas. O matplotlib é uma biblioteca abrangente para criar visualizações estáticas, animadas e interativas em Python. Já o pacote seaborn consiste numa biblioteca para fazer gráficos estatísticos em Python. Portanto, fornece uma interface de alto nível para matplotlib e integra-se com as estruturas de dados pandas. As funções da biblioteca seaborn expõem gráficos estatísticos esteticamente mais apelativos e interessantes.

Assim sendo, foram criados histogramas (sns.histplot), caixas de bigodes (sns.boxplot), gráficos de dispersão (sns.relplot) e modelos de regressão (sns.lmplot) para análise exploratória dos dados. Cada gráfico foi ajustado às suas características, sendo de salientar que se adicionou em todos título, legendas, nome dos eixos, cores e limites dos eixos. Além disso, formatou-se o tipo de letra e o tamanho de igual forma para todos os gráficos.

Portanto, cada gráfico pode ser acedido pelo utilizar clicando no botão correspondente ao parâmetro que pretende observar, uma vez que o comando "plt.show()" permite o aparecimento do mesmo. Tal como no menu, estes botões foram cria-

dos pelo pacote tkinter e apresentam comandos específicos, bem como a mesma formatação que os do menu. É apenas de salientar que a janela disponível é uma Toplevel, ou seja, é criada uma janela em cima de todas as outras janelas. Por fim, o utilizador pode regressar ao menu, utilizando o botão "Voltar ao menu".

2.3 Tabelas

De modo a criar tabelas esteticamente agradáveis, utilizaram-se os pacotes pillow e plotly. A Python Imaging Library é o pacote de processamento de imagem para a linguagem Python. Incorpora ferramentas de processamento que ajudam a editar, criar e guardar imagens. O pillow suporta um grande número de formatos de ficheiros de imagem, incluindo BMP, PNG, JPEG e TIFF. Já a plotly é uma biblioteca de código aberto que suporta diversos tipos de gráficos. Para efetuar as tabelas foi necessário `"import plotly.graph_objects as go", "from PIL import Image, ImageTk"` e `"import io"`.

Portanto, utilizou-se o `"go.Table"` para criar a tabela e o `"go.Figure"` para que a tabela aparecesse no formato de figura. Desta forma, as tabelas foram construídas segundo a sua utilização, albergando os dados necessários. Realça-se apenas que todas as tabelas foram formatadas em termos de título, cores, tipo e tamanho de letra da mesma maneira.

Contrariamente aos gráficos, para que fosse possível a disposição das tabelas na janela, foi necessário criar uma função, que permite, através do `ImageTk.PhotoImage`, que a tabela apareça numa janela. Posto isto, cada tabela aparece após clicar no botão correspondente ao parâmetro que o utilizador pretende observar.

Esta janela (Toplevel) apresenta botões tal como já foram anteriormente descritos. Estes botões apresentam o mesmo tipo de formatação que os do menu. Por fim, o utilizador pode regressar ao menu, usando o botão "Voltar ao menu".

2.4 Introdução de Dados pelo Utilizador

Com recurso ao módulo tkinter, criou-se uma janela (Toplevel) que permite introdução de dados pelo utilizador, através do Widget de entrada `Entry()`. Ademais, foram criadas labels com os nomes das variáveis disponíveis para avaliação. Tal como nas restantes, o utilizador pode regressar ao menu, usando o botão "Voltar ao menu".

Para avaliação dos resultados gerou-se uma função capaz de correlacionar os dados introduzidos com intervalos de referência. Posta esta correlação, é apresentado um texto ao utilizador que mostra quais os valores dentro e fora dos intervalos de referência. Este texto é, posteriormente, guardado num ficheiro de

formato PDF (com o auxílio do pacote reportlab e "from reportlab.lib.pagesizes import letter", "from reportlab.pdfgen import canvas" e "from reportlab.lib.utils import ImageReader").

Portanto, no final, o utilizador dispõe de um PDF com a sua avaliação clínica e um esquema ilustrativo da etiologia da PCOS. Salienta-se, apenas, que o utilizador pode inserir valores quantas vezes quiser, sendo sempre eliminada a versão anterior do PDF e do texto que aparece na janela.

3 Limitações

Ao longo do trabalho foram encontradas diversas adversidades que, de um modo geral, foram resolvidas com sucesso através de pesquisa e trabalho em equipa. Porém, algumas das soluções encontradas não foram as mais apropriadas ou devidamente optimizadas. No caso dos botões presentes na interface, seria mais interessante o tamanho destes ser ajustado automaticamente com botões responsivos. Noutra perspetiva, não se discorreu uma forma útil e contextualizada de adicionar ao menu a possibilidade de realizar a transferência de um documento (ou vários) com dados relevantes a serem adquiridos pelo utilizador.