**Apresentação**

Boa tarde a todos. Vamos então começar a nossa apresentação sobre uma base de dados relacionada com a Diabetes.

**Introdução**

Vamos começar por uma introdução teórica sucinta, importante para perceber a correlação entre as variáveis. A diabetes mellitus é uma doença na qual o organismo não produz uma quantidade suficiente de insulina ou não responde normalmente à insulina, fazendo com que o nível de açúcar (glicose) no sangue fique excecionalmente elevado.

A insulina, uma hormona secretada pelo pâncreas, controla a quantidade de glicose no sangue, sendo que esta ajuda a transportar a glicose do sangue para dentro das células.

No diabetes tipo 1, o sistema imunológico do organismo ataca as células do pâncreas que produzem insulina e mais de 90% delas são destruídas permanentemente, pelo que o organismo deixa de produzir insulina.

No diabetes tipo 2, o organismo cria resistência aos efeitos da insulina e, assim, a insulina existente não é suficiente para atender às necessidades do organismo.

Para além disso, existe ainda outra condição, a diabetes gestacional, que tal como o nome indica, pode surgir durante a gravidez.

**Base de dados**

A base de dados utilizada está disponível no Kaggle (<https://www.kaggle.com/datasets/mathchi/diabetes-data-set>), uma base de dados com o objetivo de prever se um paciente tem diabetes com base nos meios de diagnóstico que compõem as diferentes variáveis. Esta base de dados é composta por 768 pacientes mulheres, de pelo menos 21 anos de idade e com um background genético da linhagem “Pima Indian”.

**Introdução variáveis**

1. Number of times pregnant:

A primeira variável apresentada é o número de gravidezes de cada paciente, uma vez que este pode levar ao diagnóstico de diabetes gestacional, sendo por isso, um fator de risco para o desenvolvimento de diabetes.

1. Plasma glucose concentration a 2 hours in an oral glucose tolerance test:

Concentração de glicose no sangue, utilizada como meio de diagnóstico.

1. Glycemia Values:

Variável criada por nós de forma a categorizar os indivíduos de acordo com o seu valor de concentração de glucose no sangue.

1. Diastolic blood pressure (mm Hg)

?

1. Triceps skin fold thickness (mm)

A medição da grossura da prega de pele do tricep é outra maneira de avaliar o estado nutricional do paciente para além do BMI.

1. 2-Hour serum insulin (mu U/ml)

Concentração de insulina fisiológica no sangue, no período pós-prandial.

1. Body mass index (weight in kg/(height in m)^2):

A obesidade é o principal fator de risco para o desenvolvimento do diabetes tipo 2 e 80% a 90% das pessoas com este distúrbio estão acima do peso ou são obesas.

1. Diabetes pedigree function:

Uma função que pontua a probabilidade de diabetes com base na história familiar.

1. Age (years):

Antigamente, o diabetes tipo 2 era raro em crianças e adolescentes, mas vem se tornando mais comum. Geralmente começa em pessoas com idade acima de 30 anos e se torna progressivamente mais comum com o avanço da idade. Tendo em conta que 90 a 95% dos casos correspondem a diabetes tipo 2, a idade torna-se um grande fator de risco, uma vez que aproximadamente 26% das pessoas com mais de 65 anos têm diabetes tipo 2.

1. Class variable (0 or 1):

Distribuição de classes: (o valor de classe 1 é interpretado como "testado positivo para diabetes")

**Objetivo**

Identificar as variáveis e as respetivas correlações com o outcome de doença, através de análise de dados com o software Python, integrando, ainda, um menu interativo e outras funcionalidades.

**Metodologia**