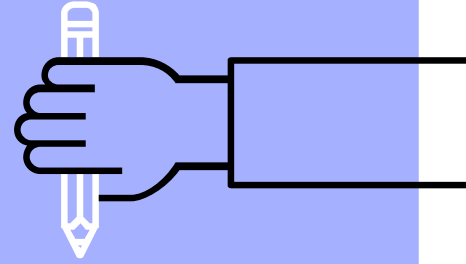
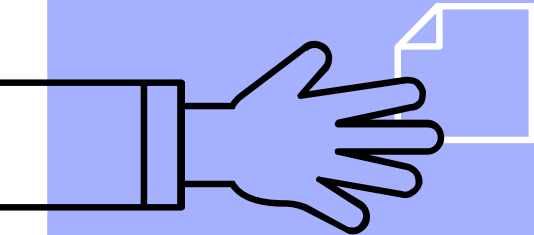


PREVISÃO DE RISCO DE DIABETES EM ESTÁGIO INICIAL

João Pedro de Lima Pereira

1. O PROBLEMA



O PROBLEMA

A **diabetes** é uma doença crônica que ocorre quando o pâncreas não consegue produzir a **insulina** ou quando o corpo não consegue utilizar de maneira adequada a insulina produzida por ele.



O PROBLEMA

Existem três tipos principais de diabetes:

Tipo 1

É causada pela reação autoimune em que o sistema imunológico do corpo ataca as células beta produtoras de insulina.

Assim o corpo produz pouca ou nenhuma insulina.

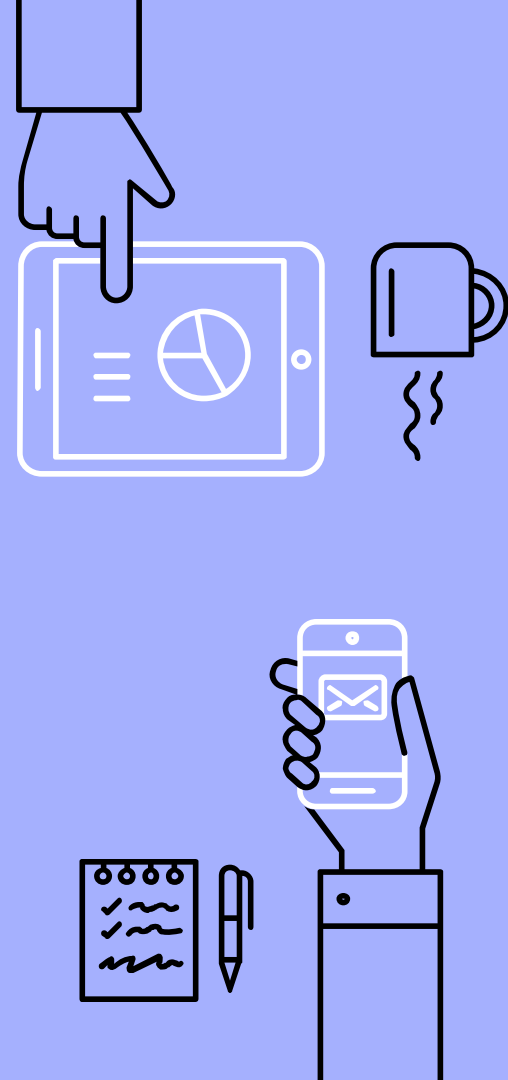
Tipo 2

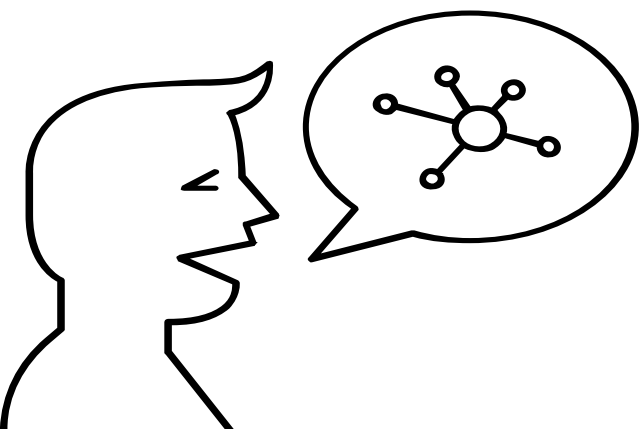
É o tipo mais comum de diabetes. Inicialmente, os níveis elevados de glicose no sangue são resultado da incapacidade das células do corpo em responder totalmente à insulina.

Diabetes Gestacional

A DMG é caracterizada por altos níveis de glicose no sangue durante a gravidez.

Pode ocorrer a qualquer momento durante a gravidez (embora mais provavelmente após a 24ª semana) e geralmente desaparece após a gravidez.





463 milhões

de diabéticos no mundo em 2019

578 milhões

de diabéticos no mundo em 2030

700 milhões

de diabéticos no mundo em 2045

Fonte: 9ª edição do Atlas da Diabetes da International Diabetes Federation (IDF), 2019

50,1%

dos adultos não foram diagnosticados

Fonte: 9ª edição do Atlas da Diabetes da International Diabetes Federation (IDF), 2019

90%

de todas as pessoas diagnosticadas com diabetes
possuíam o Tipo 2

Fonte: 9ª edição do Atlas da Diabetes da International Diabetes Federation (IDF), 2019

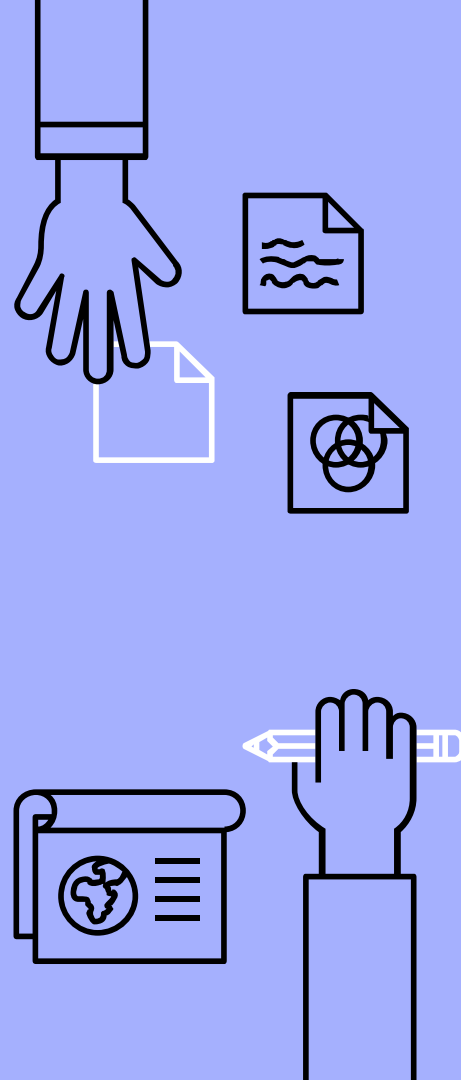
US\$ 760 bilhões

gastos com tratamento de diabetes em 2019

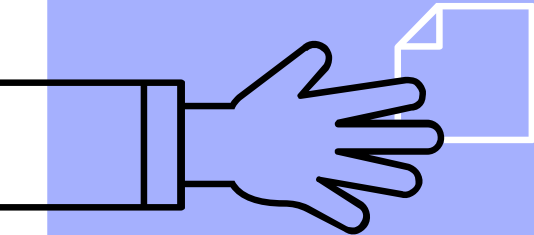
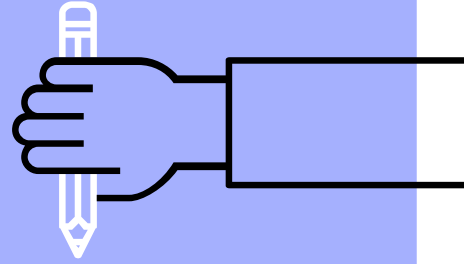
Fonte: 9ª edição do Atlas da Diabetes da International Diabetes Federation (IDF), 2019

O PROBLEMA

As evidências sugerem que o diabetes tipo 2, muitas vezes, **pode ser prevenido**, enquanto o **diagnóstico precoce** e o acesso a **cuidados adequados** para todos os tipos de diabetes **podem evitar ou retardar complicações** em pessoas que vivem com a doença.

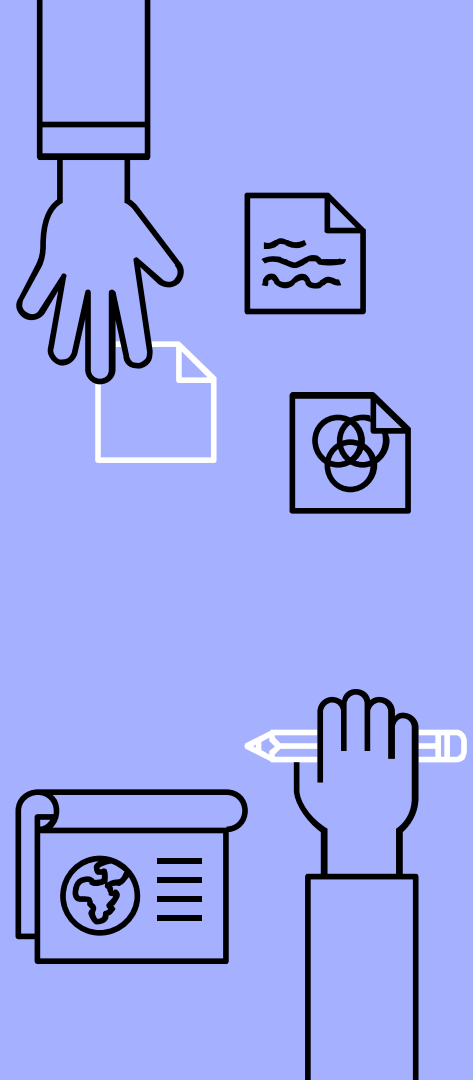


2. A Solução



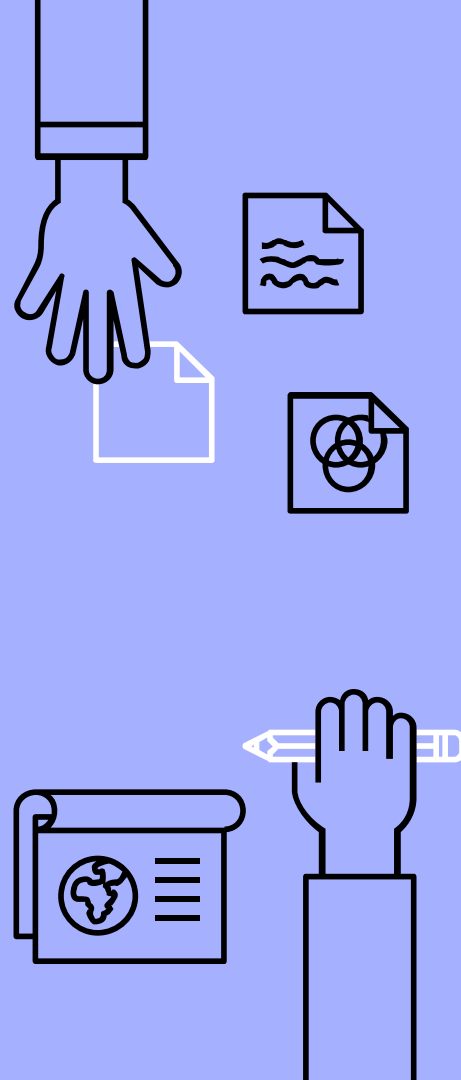
A SOLUÇÃO

Para que o diagnóstico possa ser realizado de forma que o paciente comece a receber os cuidados adequados o mais rápido possível, a solução é o desenvolvimento de um **modelo preditivo**.

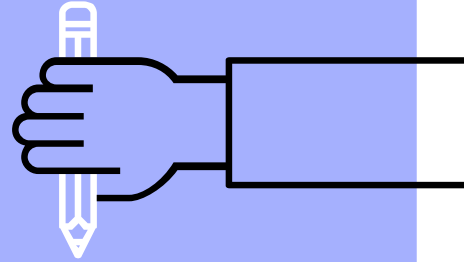
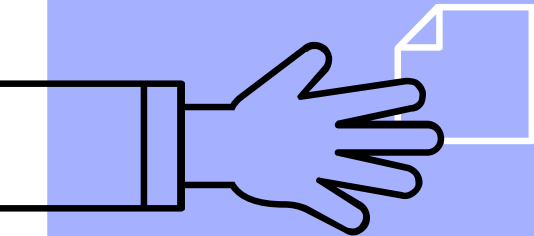


A SOLUÇÃO

Este modelo preditivo utiliza **os dados de pacientes já diagnosticados** para diagnosticar novos casos da doença de forma precoce, aumentando a chance de prevenir o diabetes tipo 2 ou tentar evitar ou retardar as complicações causadas pela doença.



3. O conjunto de dados



O CONJUNTO DE DADOS

NOME DO CONJUNTO DE DADOS

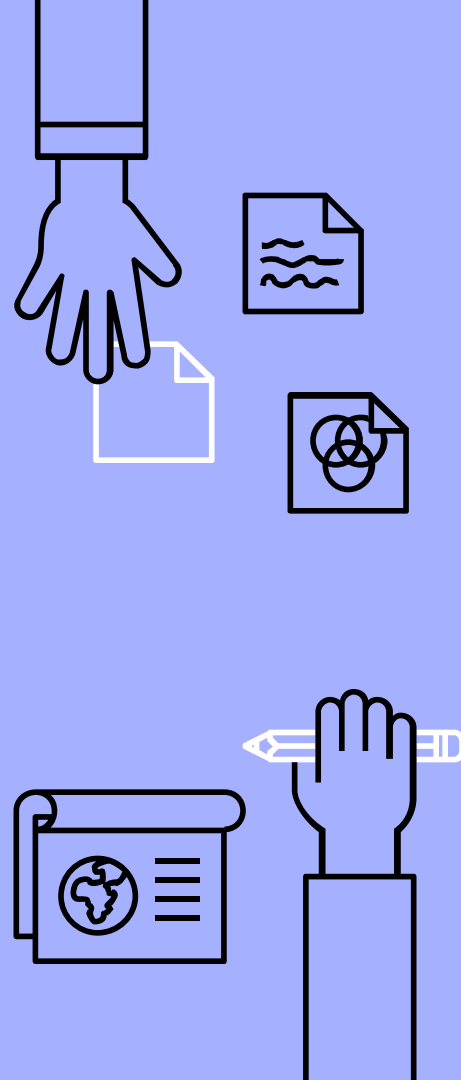
Early stage diabetes risk prediction dataset

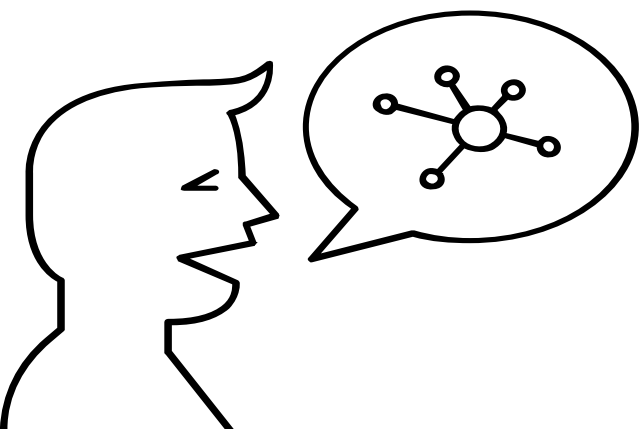
DESCRIÇÃO DO CONJUNTO DE DADOS

Ele contém dados de **diagnósticos e sintomas** de **recém diabéticos ou de pacientes que podem ser diabéticos**. Eles foram coletados usando questionários diretos de pacientes do Hospital de Diabetes de Sylhet em Bangladesh e foram aprovados por um médico.

RESPONSÁVEIS PELO CONJUNTO DE DADOS

M M Faniqul Islam (Queen Mary University of London, United Kingdom),
Rahatara Ferdousi (Metropolitan University Sylhet, Bangladesh),
Sadikur Rahman, and Humayra (Metropolitan University Sylhet, Bangladesh)
Yasmin Bushra (Metropolitan University Sylhet, Bangladesh)





520

pacientes observados

17

características observadas

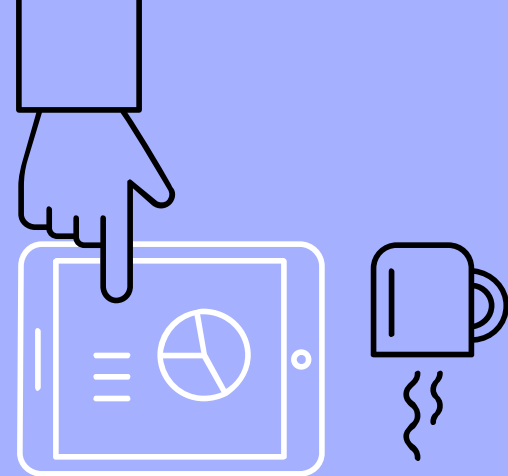
300 e 220

casos positivos e negativos,
respectivamente

O CONJUNTO DE DADOS

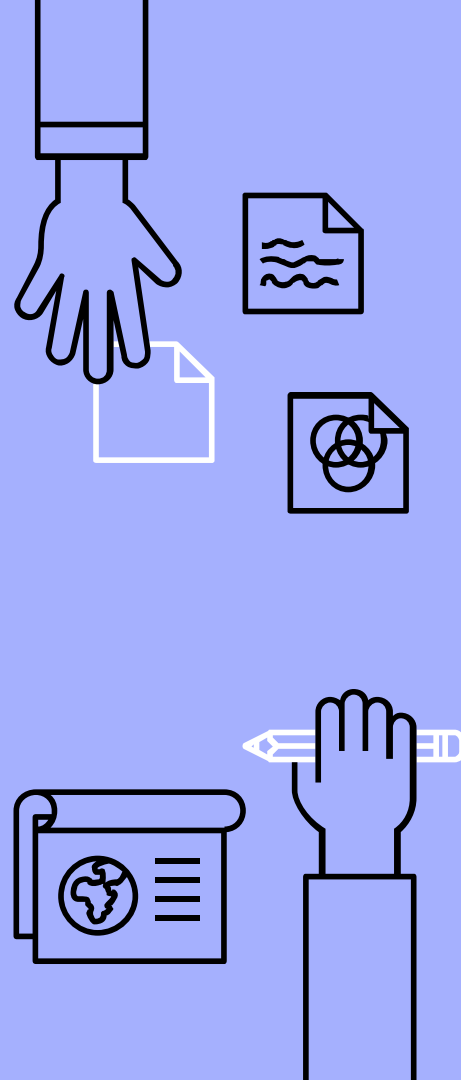
As características presentes no conjunto de dados são:

- **Idade**
- **Gênero**
- **Poliúria**
(aumento da produção de urina)
- **Polidipsia**
(sede anormal ou excessiva)
- **Perda de peso repentina**
- **Fraqueza**
- **Polifagia** (fome excessiva e ingestão anormalmente alta de sólidos pela boca)
- **Candidíase** (infecção genital por fungos)
- **Visão embaçada**
- **Coceira**
- **Irritabilidade**
- **Cicatrização demorada**
- **Paralisia parcial de algum membro**
- **Rigidez muscular**
- **Perda de pelos do corpo**
- **Obesidade**

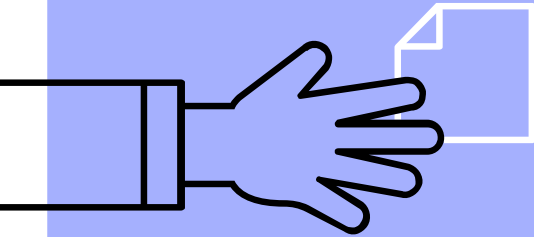
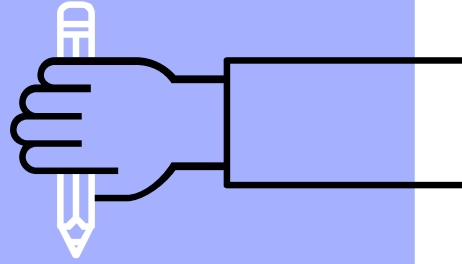


O CONJUNTO DE DADOS

Cada uma das observações dos pacientes encontradas neste conjunto de dados **possui a informação se aquele paciente foi ou não diagnosticado com diabetes**, sem a especificação exata de qual dos tipos da doença.



4. Explorando os dados

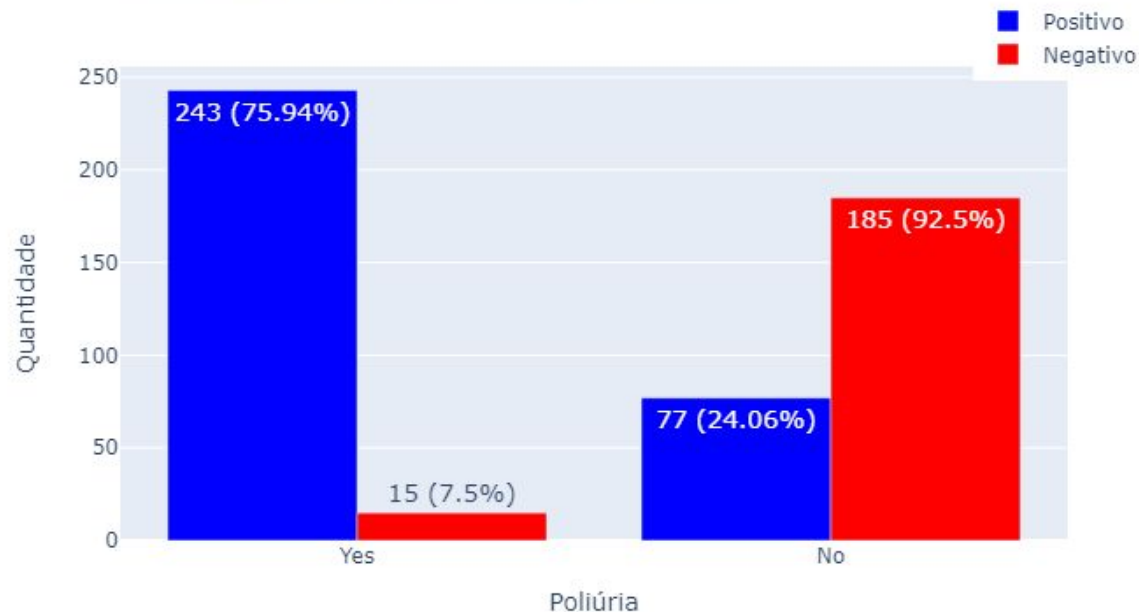


Pacientes por classe e idade

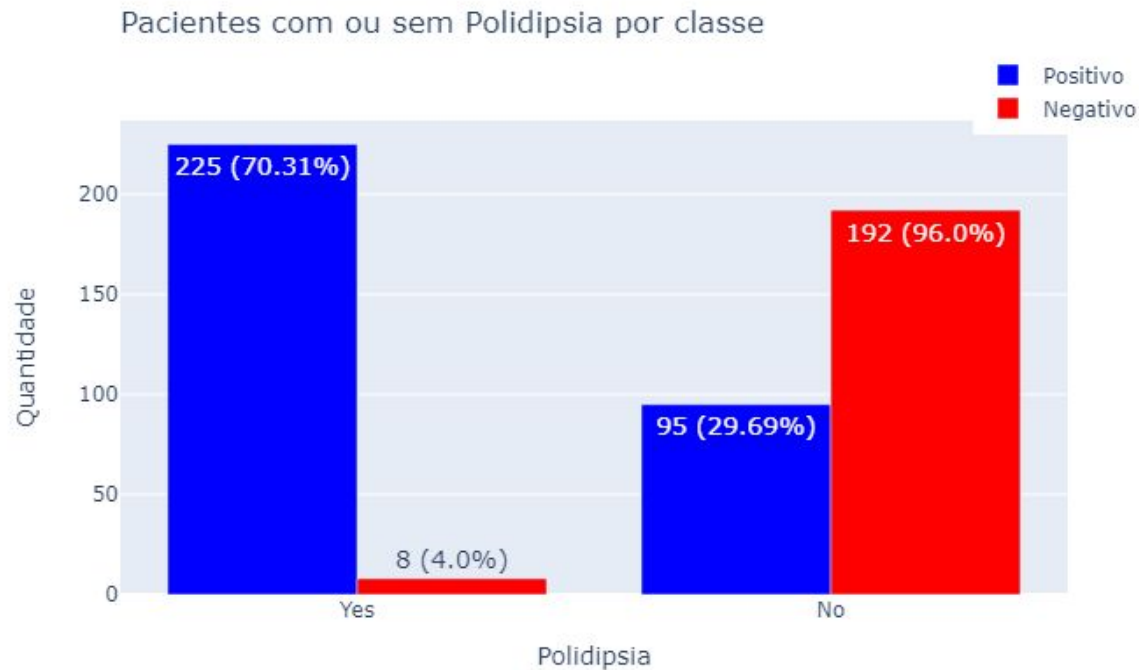


A distribuição das idades dos pacientes diagnosticados com Diabetes foi bem semelhante à dos pacientes não diagnosticados.

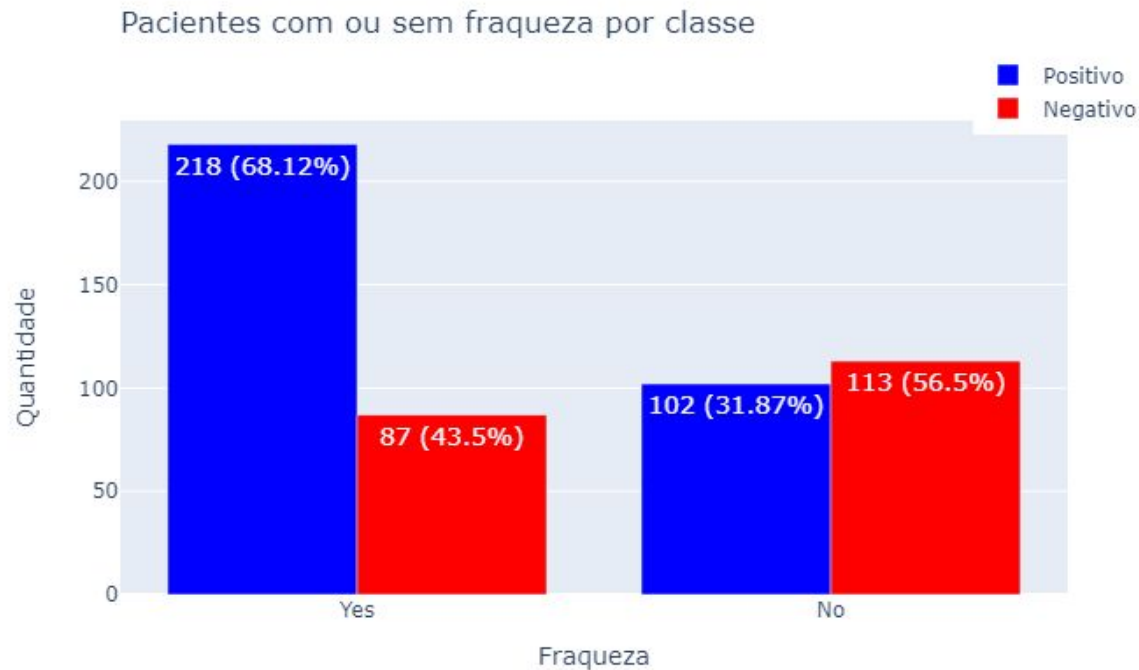
Pacientes com ou sem Poliúria por classe



A Poliúria foi característica mais presente nos pacientes diagnosticados com diabetes. Ela foi observada em quase 76% deles.

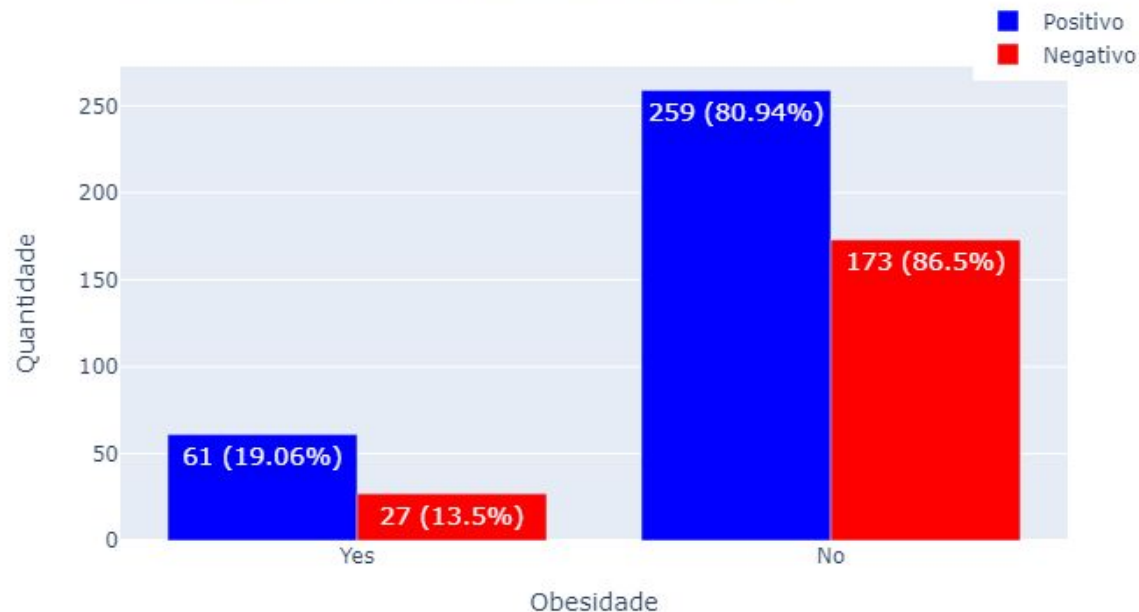


A Polidipsia foi a segunda característica mais presente nos pacientes diagnosticados com diabetes. Ela foi observada em pouco mais de 70% deles.



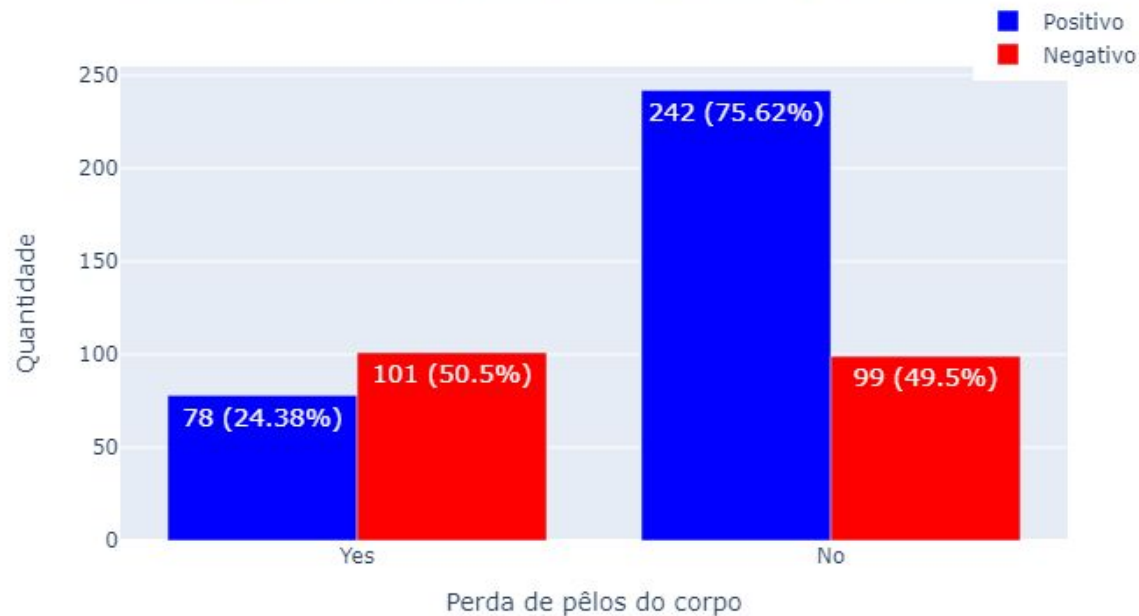
A fraqueza foi a terceira característica mais presente nos pacientes diagnosticados com diabetes. Ela foi observada em pouco mais de 68% deles.

Pacientes com ou sem obesidade por classe

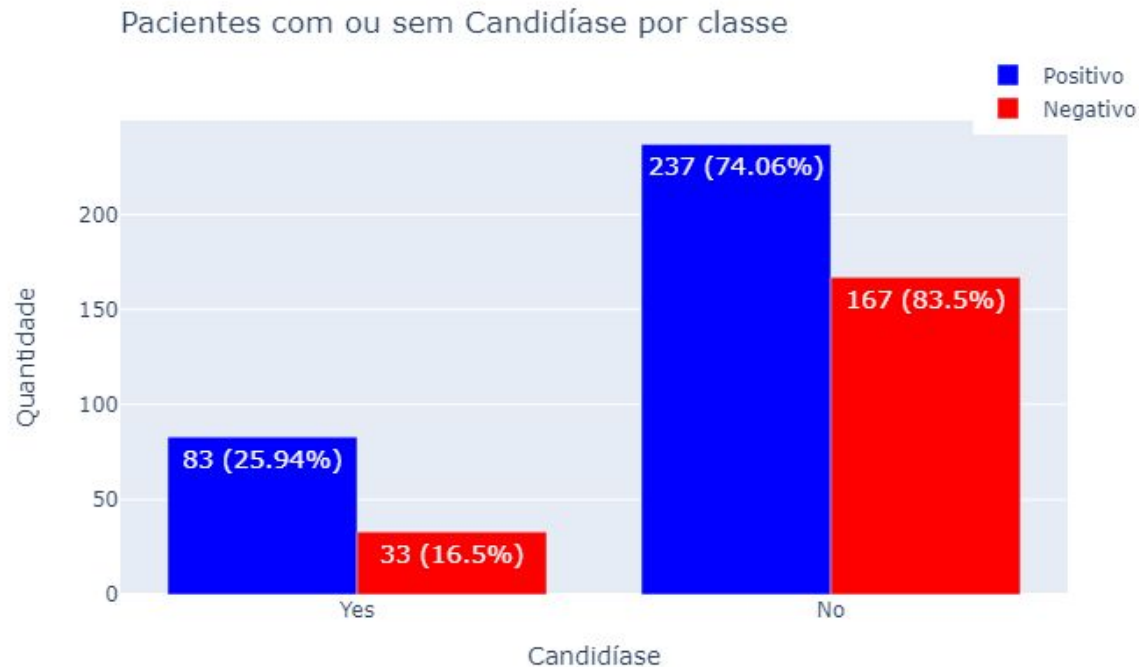


A obesidade foi a característica menos presente nos pacientes diagnosticados com diabetes. Ela foi observada apenas em pouco mais de 19% deles.

Pacientes com ou sem perda de pêlos do corpo por classe

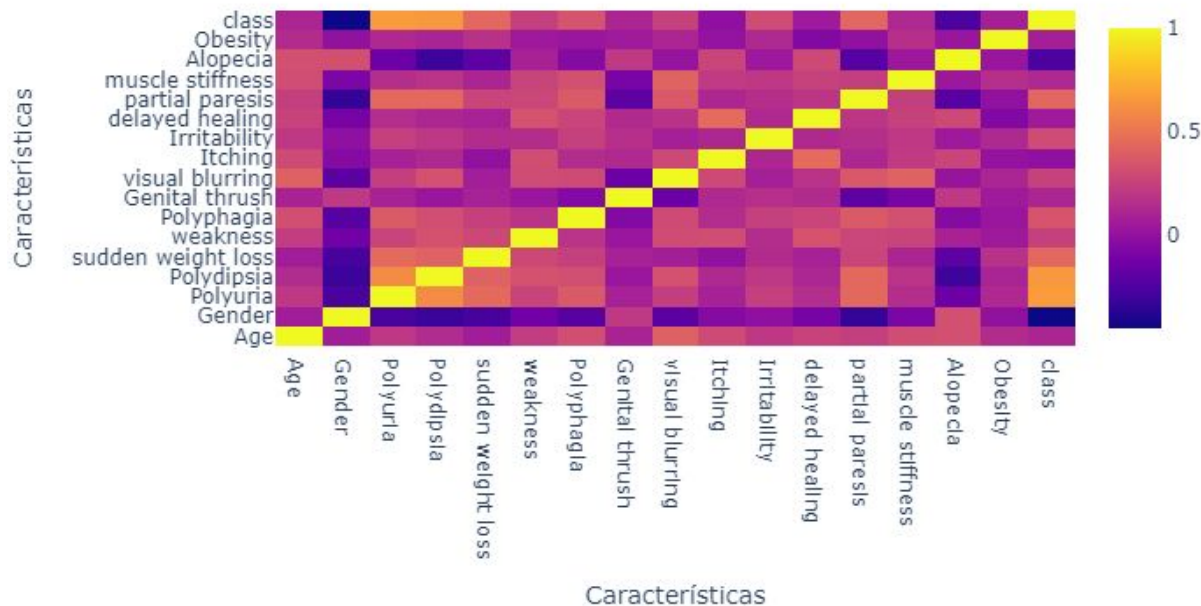


A perda de pêlos do corpo foi a segunda característica menos presente nos pacientes diagnosticados com diabetes. Ela foi observada apenas em pouco mais de 24% deles.



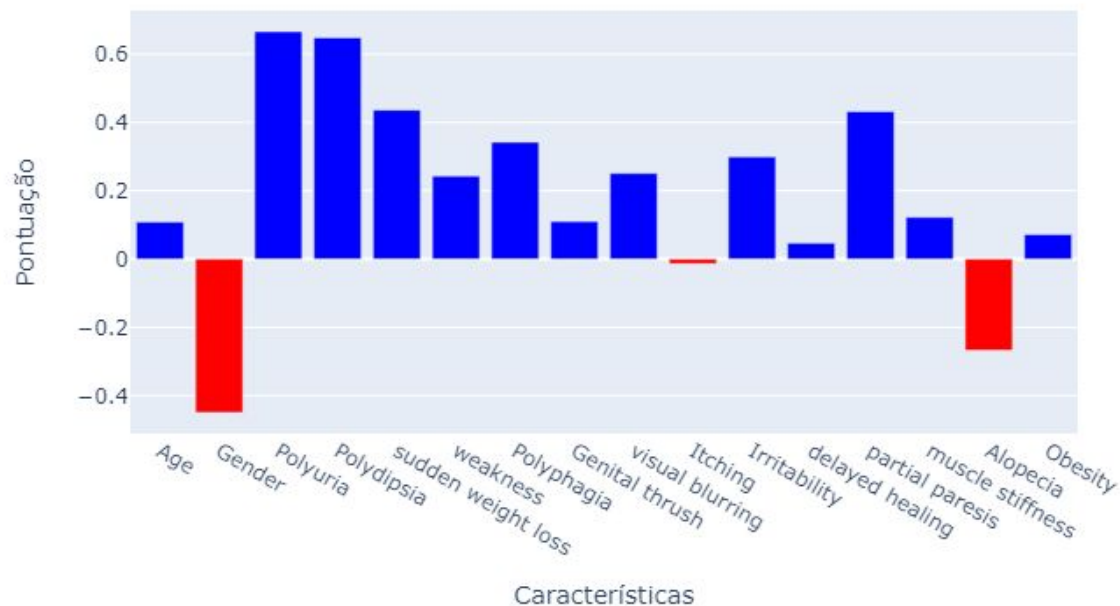
A Candidíase foi a terceira característica menos presente nos pacientes diagnosticados com diabetes. Ela foi observada apenas em quase 26% deles.

Correlação entre as características



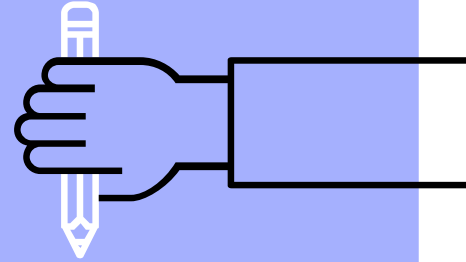
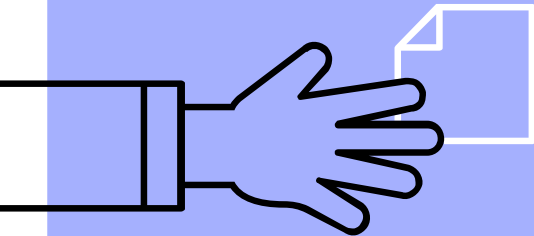
Analizando a correlação entre as características, podemos observar que a Poliúria, a Polidipsia e a perda de peso repentina são as que possuem maior correlação com o diagnóstico de Diabetes (class) - as mais claras do mapa.

Correlação entre as características e a classe



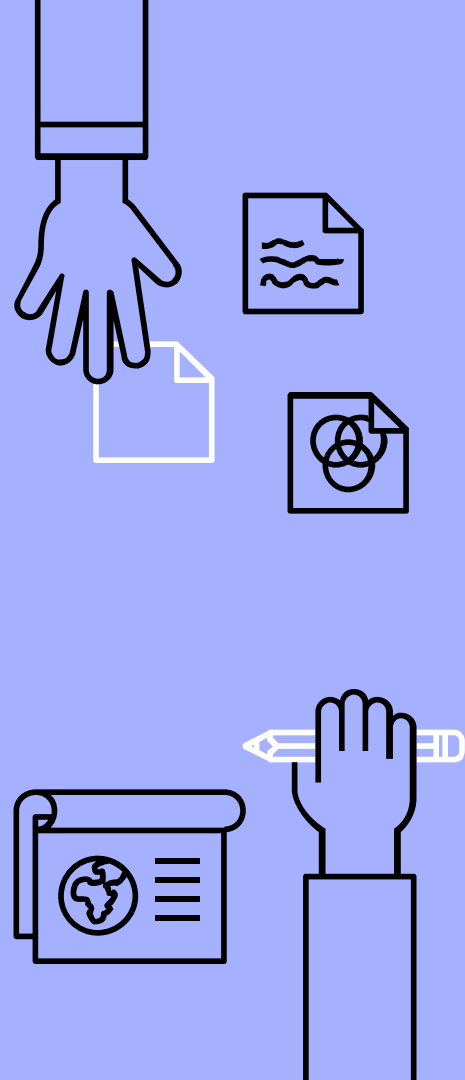
Analisando apenas os valores da correlação entre as características e a classificação do paciente como diabético ou não, podemos confirmar que a Poliúria, a Polidipsia e a perda de peso repentina são as que possuem maior correlação.

5. 0 modelo preditivo



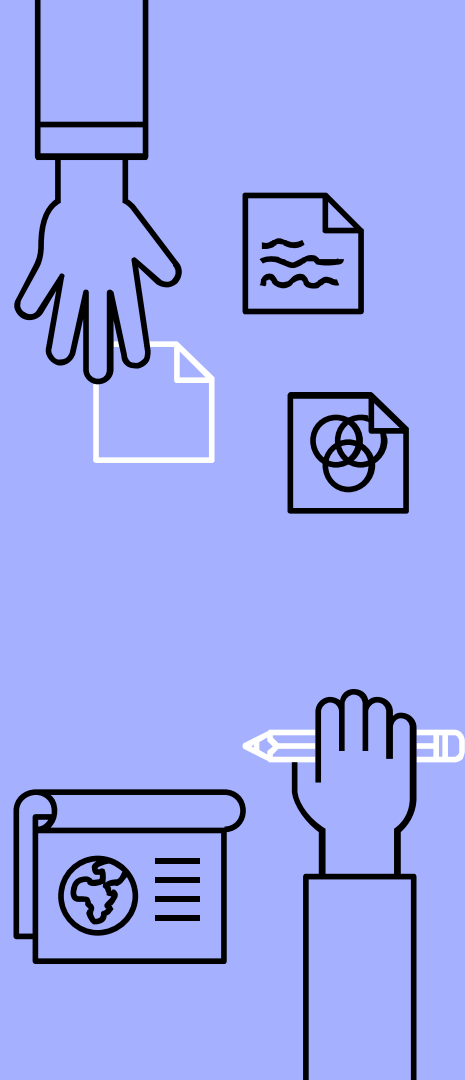
O MODELO PREDITIVO

O modelo preditivo a ser implementado para resolver este problema deve reconhecer padrões dentro do conjunto de dados e **classificar** os pacientes indicando um possível diagnóstico de diabetes ou não.



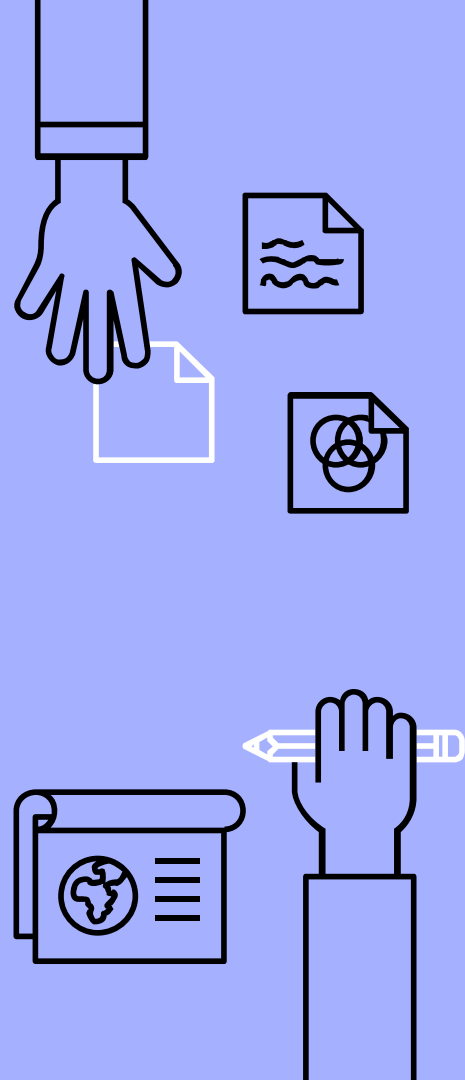
O MODELO PREDITIVO

Por se tratar de um problema de classificação de pacientes de determinada doença, deve-se procurar construir um modelo preditivo que tenha um **baixo índice de falsos negativos** - o chamado recall - mesmo que isso faça com que ele identifique alguns pacientes saudáveis como doentes.



O MODELO PREDITIVO

O mapa a seguir mostrará como foi feita a escolha dos **estimadores** a serem testados para a construção do modelo.

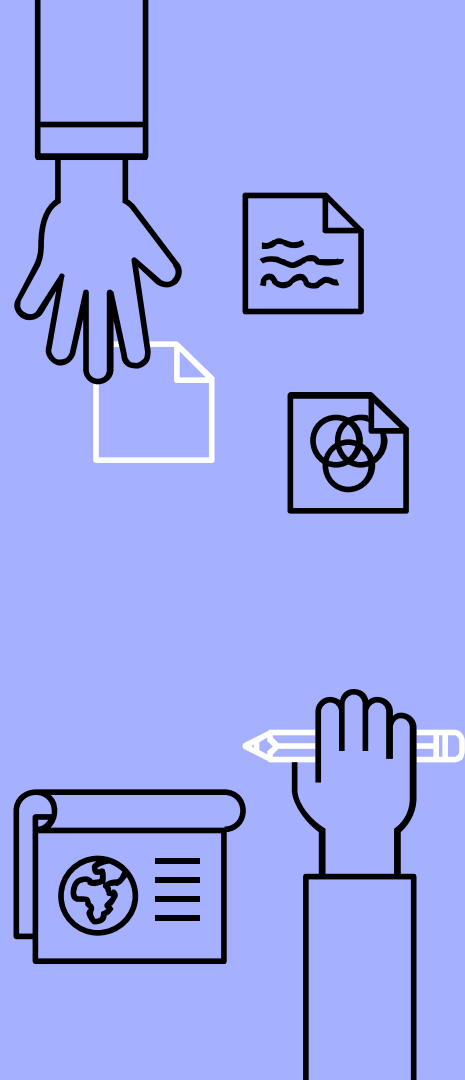


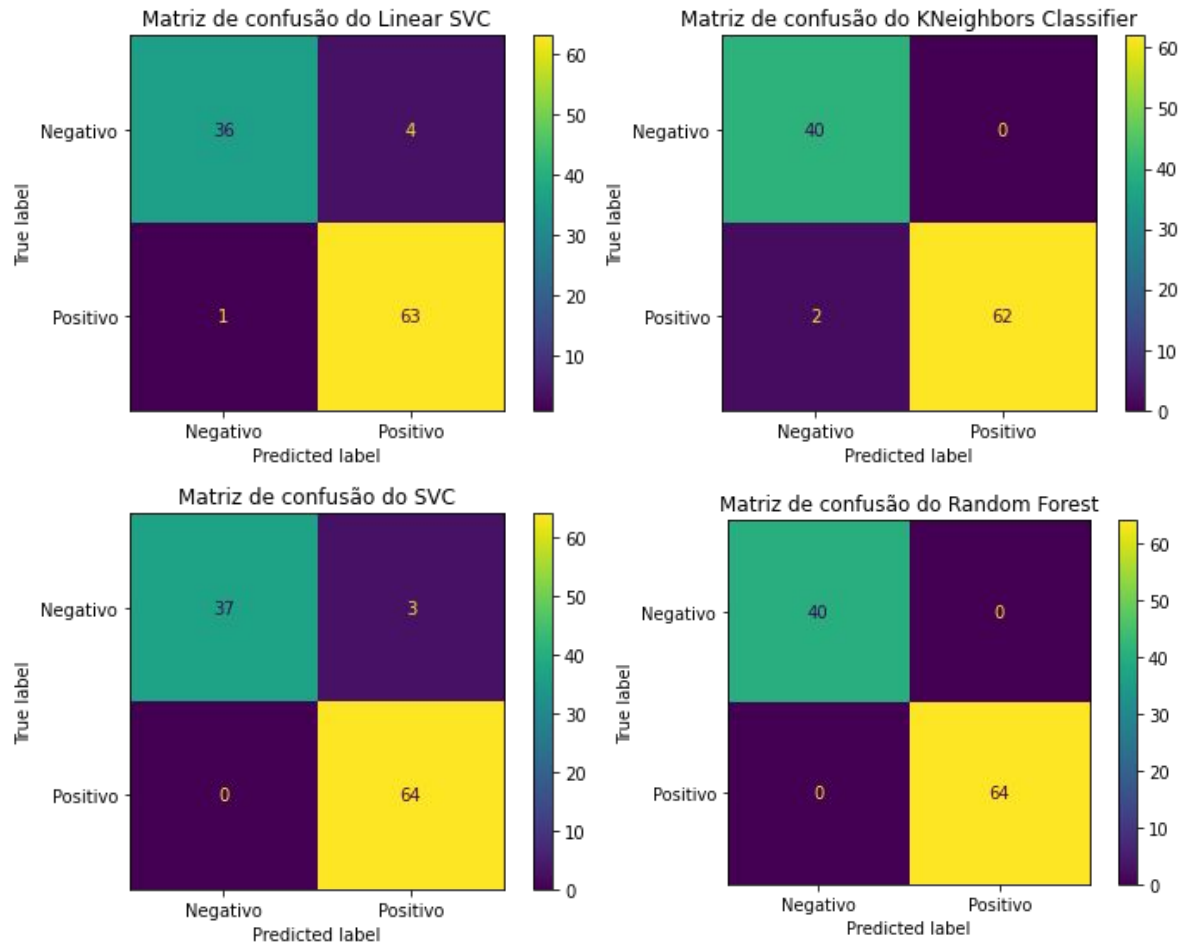
32

O MODELO PREDITIVO

O conjunto de dados é **dividido** para o **treinamento (80%)** e **teste (20%)** dos estimadores.

A ideia é que o estimador **encontre os padrões** no conjunto de treinamento e depois tente **prever corretamente** qual a classe cada observação do conjunto de teste pertence.





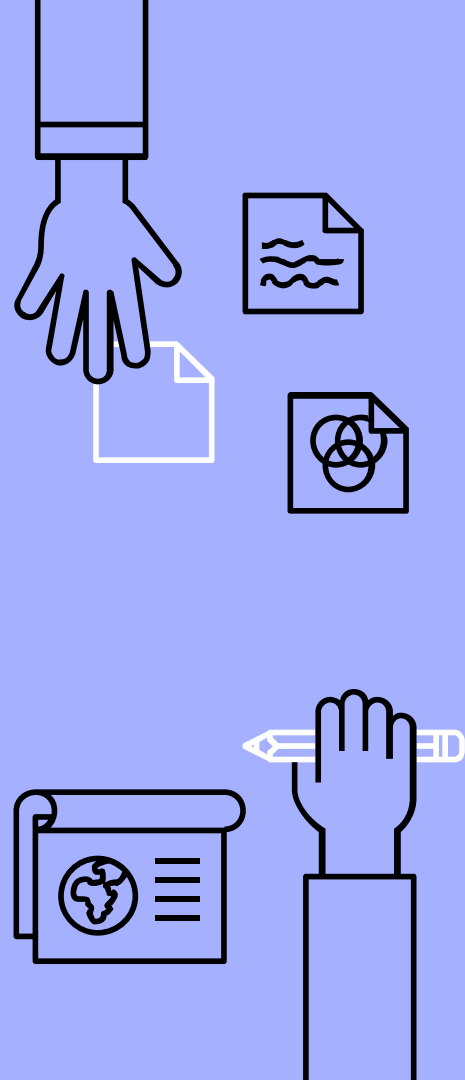
As matrizes de confusão ao lado mostram as previsões realizadas por cada um dos estimadores.

O Linear SVC errou a categoria de 5 observações, o KNN 2 observações, o SVC 3 observações, enquanto o

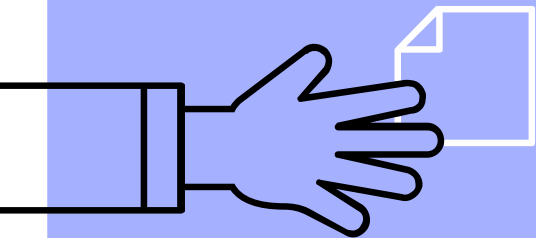
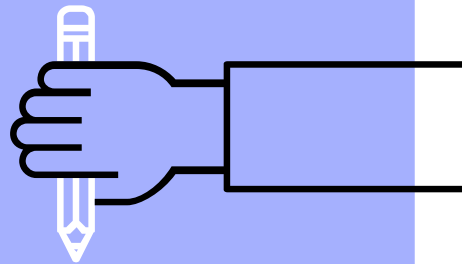
Random Forest acertou todas as previsões.

O MODELO PREDITIVO

Desta forma, o estimador escolhido para o modelo preditivo foi o **Random Forest** por ter obtido o melhor desempenho entre os quatro estimadores testados.

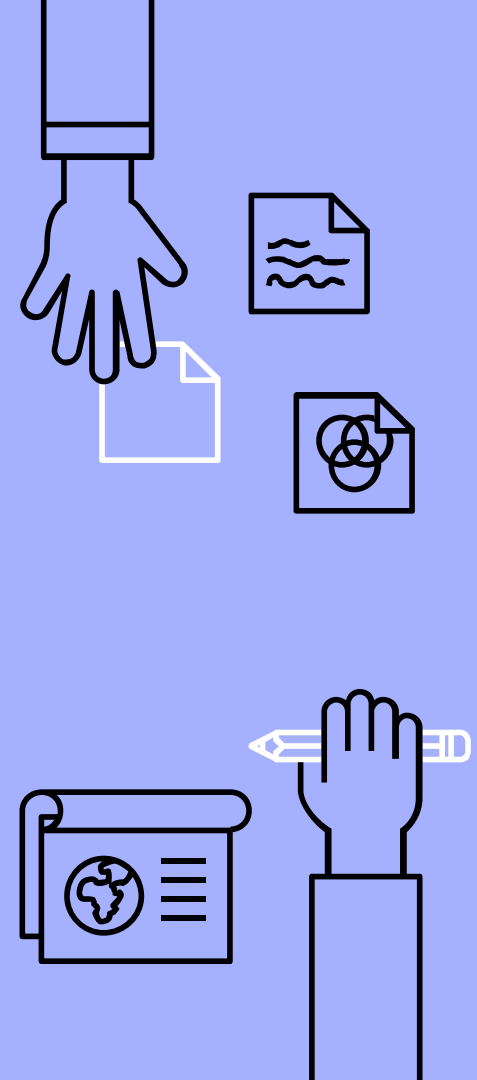


6. As questões



AS QUESTÕES

- 1) Quais os 3 principais fatores que levam a um diagnóstico positivo de diabetes?
- 2) Qual a chance de uma pessoa do sexo masculino com menos de 40 anos ter o diagnóstico positivo de diabetes?
- 3) Qual a relação entre Polyphagia e o diagnóstico de diabetes?



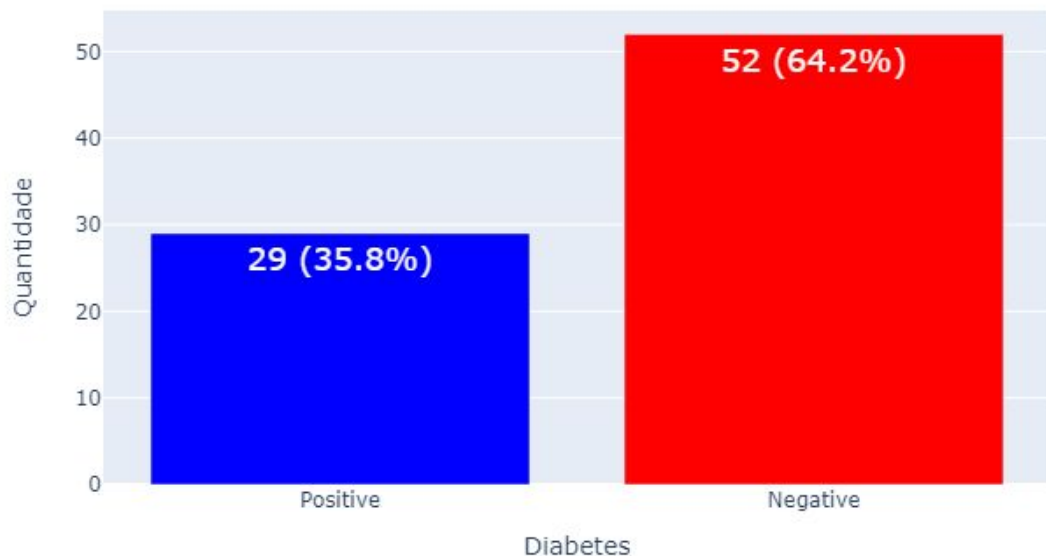
- 1) Quais os 3 principais fatores que levam a um diagnóstico positivo de diabetes?

Característica	Pontuação
Polidipsia	120,78
Poliúria	116,18
Perda de peso repentina	57,74
Paralisia parcial de algum membro	55,31
Gênero	38,74
Irritabilidade	35,33
Polifagia	33,19
...	...

Utilizando o método de seleção de características **SelectKBest**, as três características com a maior pontuação foram **Polidipsia**, **Poliúria** e **perda de peso repentina**.

- 2) Qual a chance de uma pessoa do sexo masculino com menos de 40 anos ter o diagnóstico positivo de diabetes?

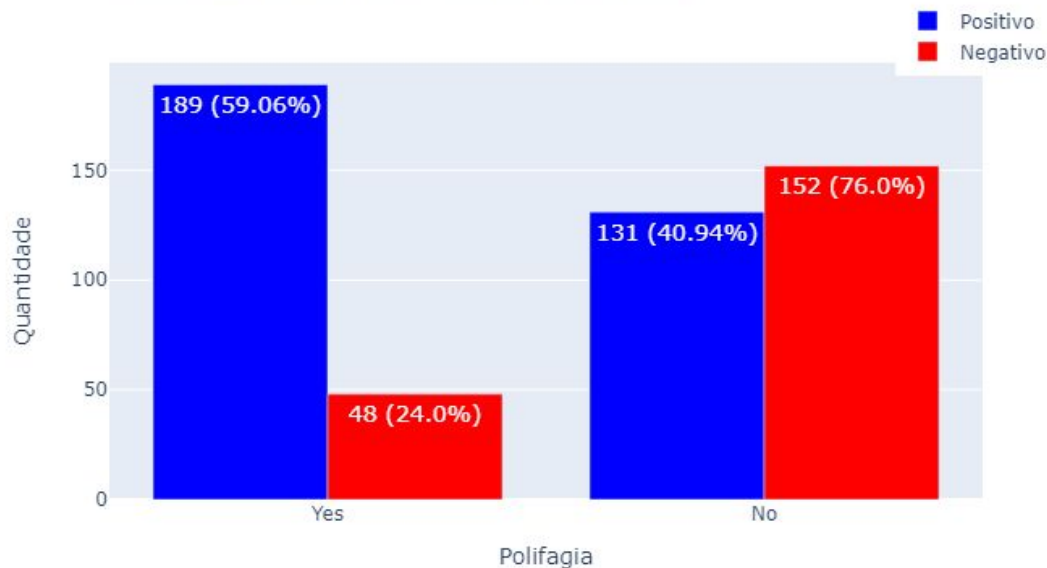
Pacientes do sexo masculino com menos de 40 anos por classe



Analizando os dados dos pacientes do sexo masculino com menos de 40 anos, temos que a chance deles serem diagnosticados com Diabetes é de **35,8%**.

3) Qual a relação entre Polyphagia e o diagnóstico de diabetes?

Pacientes com ou sem Polifagia por classe



De acordo com o conjunto de dados, **59,06%** dos pacientes com diagnóstico de Polifagia (Polyphagia) também foram diagnosticados com Diabetes.

OBRIGADO!

Alguma dúvida?



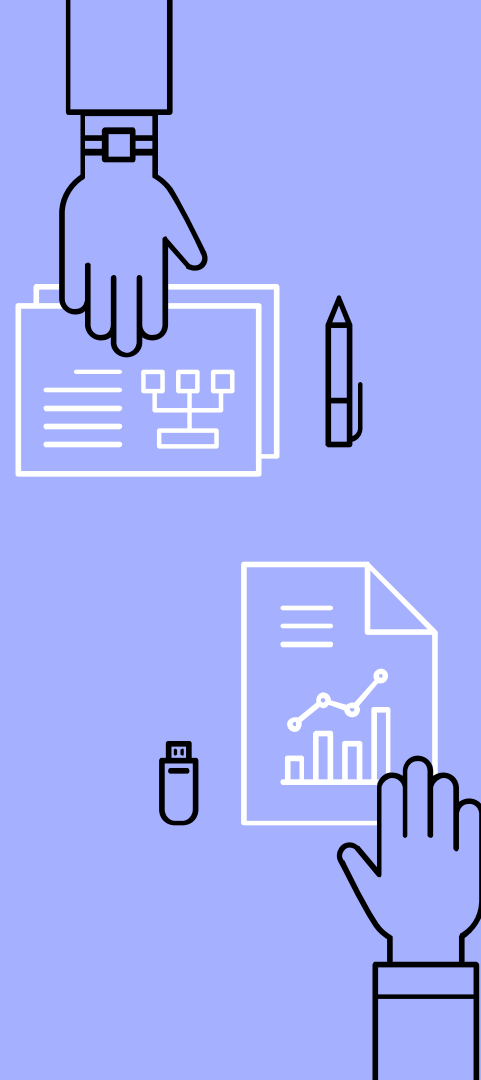
ONDE ENCONTRAR

Todo a documentação deste trabalho pode ser encontrada no repositório do GitHub:

<https://github.com/jplpereira/early-stage-diabetes>

O trabalho com o conjunto de dados encontra-se especificamente no notebook:

<https://github.com/jplpereira/early-stage-diabetes/blob/master/Diabetes.ipynb>



O modelo desta apresentação foi disponibilizado gratuitamente em



Para PowerPoint e
Google Slides



100% gratuitos para
uso pessoal ou
comercial



Prontos para uso,
profissional e
customizável



Surpreenda seu
público com visuais
atraentes