UNIVERSIDADE FEDERAL DO CEARÁ

TECHNICAL REPORT

Aluno: Guilherme Pinheiro Serafim

1. Introdução

Este relatório técnico aborda a análise do dataset "Predict Diabetes" ou "Diagnóstico de Diabetes", explorando suas características principais e objetivos de análise.

Este dataset contém informações sobre pacientes e a presença ou ausência de diabetes (coluna "Outcome"). O objetivo é realizar diversas técnicas de aprendizado de máquina para classificação e agrupamento, incluindo a seleção de melhores hiperparâmetros para o KNN, a determinação do número ideal de clusters via métodos do cotovelo e silhueta, a seleção de atributos via Lasso e a análise da distribuição dos clusters em relação à variável alvo.

As variáveis incluem:

- ***Gravidezes:*** *Número de vezes que a paciente esteve grávida.*
- ***Glicose:*** *Níveis médios de glicose no sangue.*
- ***Pressão Arterial:*** *Valor médio da pressão sanguínea.*
- ***Espessura da Pele:*** *Medida da dobra cutânea em milímetros.*
- ***Insulina:*** *Níveis de insulina presentes no sangue.*
- ***Índice de Massa Corporal (IMC):*** *Relação entre peso e altura, indicador de obesidade.*
- ***Histórico Genético:*** *Pontuação que reflete a predisposição genética ao diabetes.*
- ***Idade:*** *Idade dos pacientes.*
- ***Diagnóstico:*** *Variável binária indicando presença (1) ou ausência (0) de diabetes.*

Esta análise pretende:

- *Explorar a relação entre os fatores acima e o diagnóstico de diabetes.*
- *Construir um modelo preditivo que permita identificar pacientes em risco.*

2. Observações

Nenhum problema ou imprevisto relevante ocorreu durante o processo de análise.

3. Resultados e discussão

***Questão 1** - O objetivo desta questão é aplicar o modelo de K-Nearest Neighbors (KNN) utilizando o GridSearchCV da biblioteca scikit-learn para encontrar a melhor configuração de parâmetros para o modelo de classificação. A análise envolve a escolha da métrica de distância, o número de vizinhos e os pesos utilizados pelo classificador.*

Etapas do Fluxograma do Processo:

- ***Carregamento do Dataset***

O dataset "diabetes.csv" foi carregado usando a biblioteca pandas. A variável `x` contém as features, enquanto `y` é o target, a coluna "Outcome" que indica a presença ou ausência de diabetes.

- ***Pré-processamento dos Dados***

Divisão dos dados em treino (80%) e teste (20%) usando train_test_split.

Normalização dos dados usando StandardScaler, para garantir que as variáveis estejam na mesma escala e, assim, melhorar o desempenho do modelo.

- ***Definição de Parâmetros para GridSearchCV***

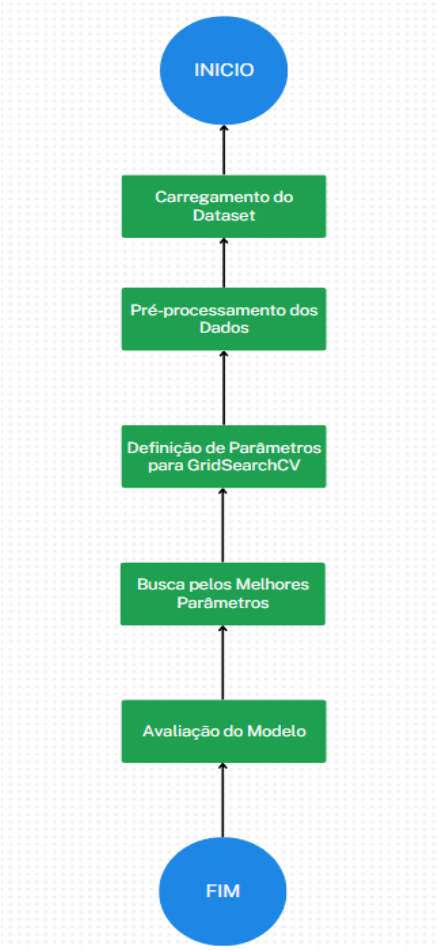
1. *O GridSearchCV foi configurado para testar diferentes combinações de parâmetros:*
2. *n_neighbors: número de vizinhos, variando de 1 a 20.*
3. *metric: métrica de distância, que pode ser `euclidean`, `manhattan` ou `minkowski`.*
4. *weights: pesos a serem atribuídos aos vizinhos, podendo ser `uniform` ou `distance`.*

- ***Busca pelos Melhores Parâmetros***

O GridSearchCV realiza uma busca exaustiva por todos os parâmetros possíveis, utilizando validação cruzada de 5 folds (`cv=5`) e buscando maximizar a acurácia.

- ***Avaliação do Modelo***

Após a execução do GridSearchCV, o modelo com os melhores parâmetros foi selecionado. A acurácia do modelo no conjunto de teste foi avaliada, juntamente com o relatório de classificação, que detalha as métricas de precisão, recall e F1-score para cada classe.



[Figura 1 - Fluxograma do Processo].

- ***Melhores Parâmetros Encontrados**:*

1. ***n_neighbors**: 15*
2. ***metric**: Manhattan*
3. ***weights**: Uniform*

- ***Acurácia no Conjunto de Teste**: 77.27%*

- ***Relatório de Classificação**:*

Classe	Precisão	Recall	F1-Score	Suporte
0	0.79	0.88	0.83	100
1	0.72	0.57	0.64	54

[Figura 2 - Tabela do Relatório de Classificação].

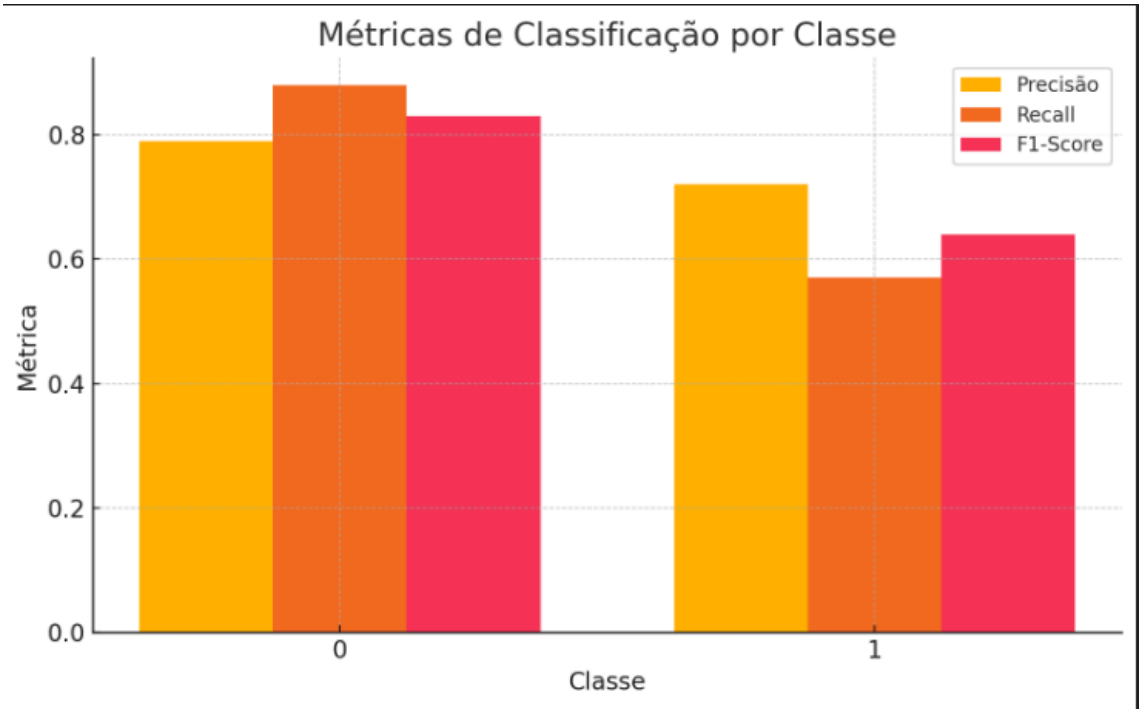
Métricas Globais:

- ***Média Macro:***

1. *Precisão: 0.76*
2. *Recall: 0.73*
3. *F1-Score: 0.74*

- ***Média Ponderada:***

1. *Precisão: 0.77*
2. *Recall: 0.77*
3. *F1-Score: 0.77*



[Figura 3 - Métricas de Classificação por Classe].

- ***Análise dos Resultados:***

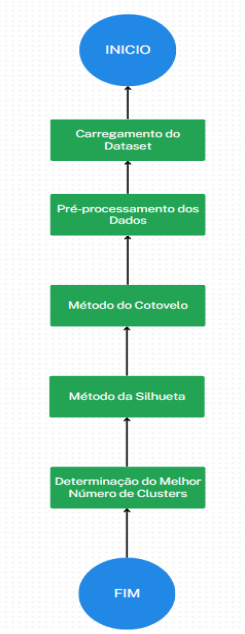
A análise dos resultados mostra que o modelo apresenta um bom desempenho geral, com uma acurácia de 77.27%. A classe 0, que representa os indivíduos sem diabetes, obteve uma precisão e recall mais altos, o que é esperado devido à maior quantidade de amostras dessa classe no dataset. A classe 1, com menor quantidade de amostras, obteve desempenho inferior, refletido em uma precisão e recall mais baixos. Comparando com os resultados obtidos na AV1, o GridSearchCV proporcionou uma configuração de parâmetros mais robusta, indicando que o ajuste fino dos hiperparâmetros pode melhorar o desempenho do modelo em relação à análise inicial.

O uso do GridSearchCV se mostrou eficaz na busca pelos melhores parâmetros para o modelo KNN. A aplicação de validação cruzada e a escolha adequada dos parâmetros ajudaram a otimizar o desempenho do modelo. O próximo passo seria explorar outras técnicas de balanceamento de classes, como o uso de técnicas de sobremuestreamento ou subamostragem, para tentar melhorar o desempenho na classe minoritária.

Questão 2 *- O objetivo desta questão é determinar o número ideal de clusters para o modelo K-Means, sem considerar a coluna "Outcome" (target). Para isso, foram utilizados dois métodos: Método do Cotovelo e Método da Silhueta, que ajudam a identificar o valor de kkk que melhor divide os dados em clusters.*

Etapas do Fluxograma do Processo:

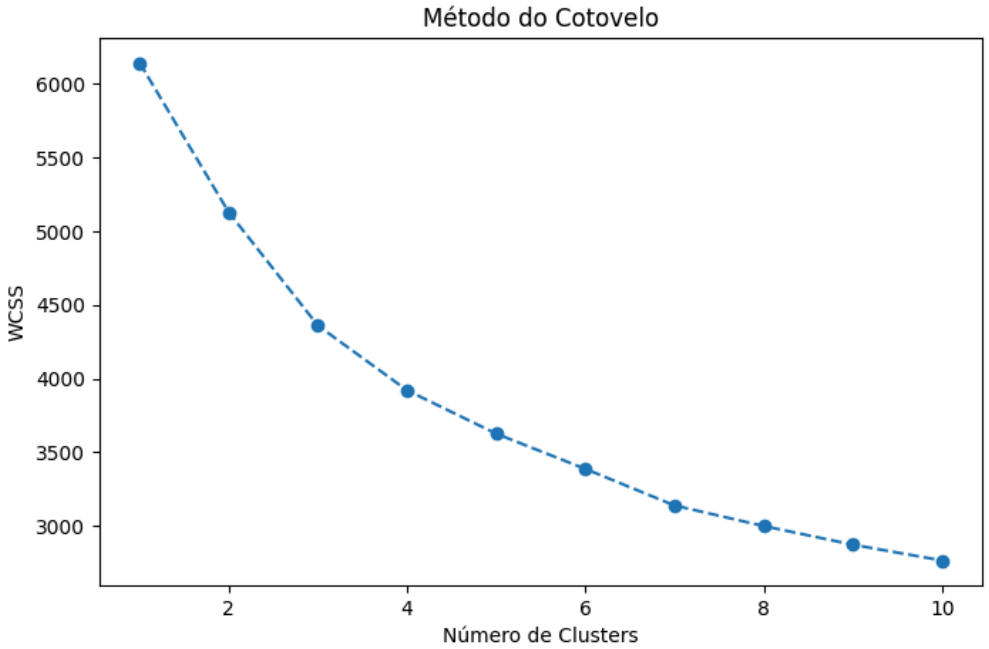
- ***Carregamento do Dataset:*** *O dataset "diabetes.csv" foi carregado, e a variável* `X` *foi obtida ao remover a coluna "Outcome", enquanto* `y` *representa a coluna alvo.*
- ***Pré-processamento dos Dados:*** *A normalização dos dados foi realizada utilizando o* `StandardScaler`*, para garantir que todas as variáveis estivessem na mesma escala.*
- ***Método do Cotovelo:*** *O método do cotovelo foi utilizado para identificar o valor de* ***k*** *onde ocorre uma redução mais significativa na soma dos quadrados intra-cluster (WCSS). A análise foi feita variando o número de clusters de 1 a 10.*
- ***Método da Silhueta:*** *O índice de silhueta foi calculado para cada valor de* ***k*** *de 2 a 10. Esse método avalia o quão bem os dados estão agrupados. O valor de* ***k*** *que maximiza o índice de silhueta é considerado o melhor número de clusters.*
- ***Determinação do Melhor Número de Clusters:*** *Os dois métodos forneceram sugestões sobre o melhor valor de* ***k****, que foi comparado para identificar a escolha ideal.*



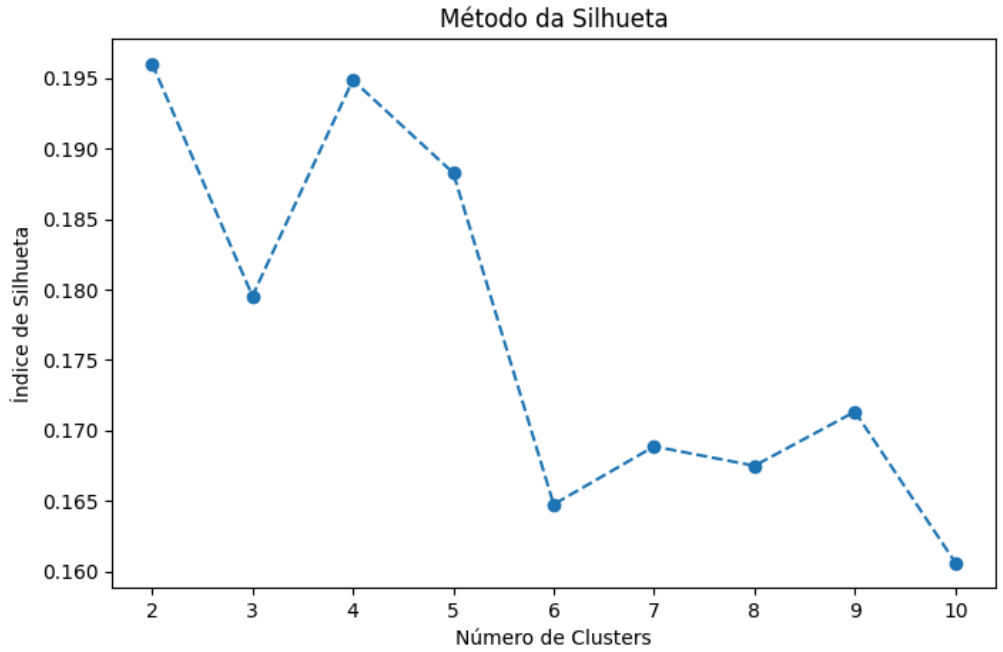
[Figura 4 - Fluxograma do Processo].

- ***Resultado e Análise:***

1. ***Método da Silhueta****: O índice de silhueta foi mais alto em k=2 = e k=4, com um pequeno favorecimento para k=2, indicando que esses valores de* ***k*** *podem gerar bons agrupamentos.*
2. ***Método do Cotovelo****: O gráfico do cotovelo indicou que o valor de k=4 é uma boa escolha, pois a diminuição da WCSS após esse ponto se torna menos expressiva.*
3. *O* ***Método do Cotovelo*** *sugere que k=4 é o número ideal de clusters, pois a redução na WCSS se estabiliza nesse ponto.*
4. *O* ***Método da Silhueta*** *sugere que tanto k=2 quanto k=4 podem ser boas opções, com uma leve preferência por k=2.*
5. *Ambos os métodos apontam para k=4 como uma boa escolha, o que torna este valor de* ***k*** *a melhor opção para o agrupamento dos dados.*



[Figura 5 - Método do Cotovelo].



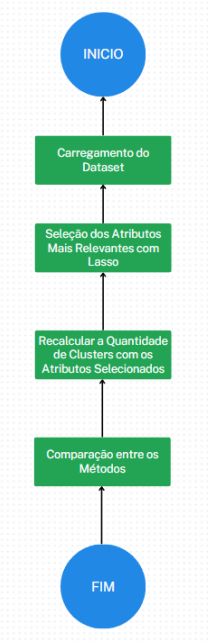
[Figura 6 - Método da Silhueta].

- *Após aplicar os dois métodos, foi identificado que o número de clusters ideal para este conjunto de dados é k=4. Ambos os métodos (cotovelo e silhueta) sugerem esse valor como o melhor para o modelo K-Means, embora o método da silhueta também tenha indicado k=2 como uma possibilidade viável.*

Questão 3 - *O objetivo dessa questão é selecionar os dois atributos mais relevantes do dataset utilizando o método de* ***Lasso****, recalcular o número ideal de clusters com os métodos do* ***Cotovelo*** *e da* ***Silhueta****, e verificar se houve alguma mudança no valor de* ***k*** *em comparação à Questão 2. Caso os métodos sugiram valores diferentes para* ***k****, scatterplots serão gerados para uma análise visual.*

Etapas do Fluxograma do Processo:

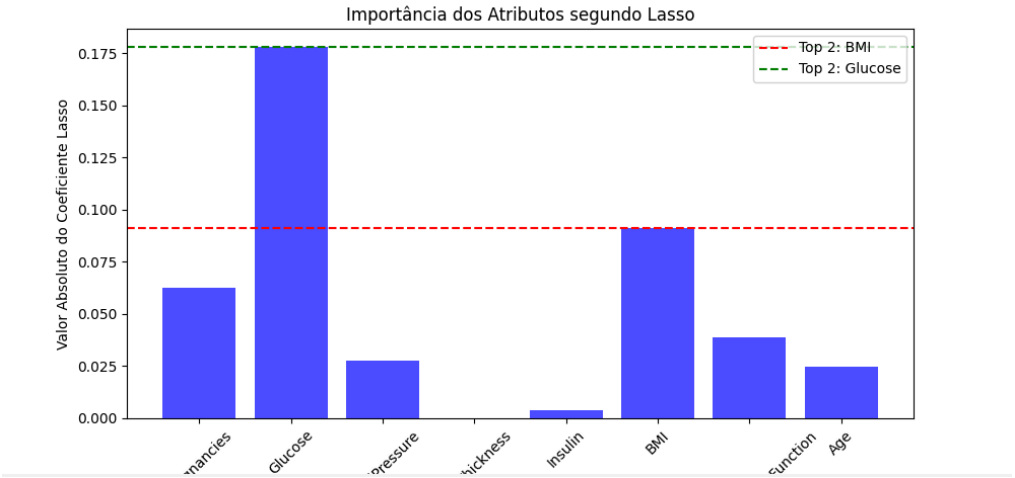
- ***Carregamento do Dataset:*** *O dataset "diabetes.csv" foi carregado e as variáveis X (features) e y (target) foram separadas. A normalização dos dados foi realizada para garantir a mesma escala para todas as variáveis.*
- ***Seleção dos Atributos Mais Relevantes com Lasso:*** *O modelo de regressão* ***Lasso*** *foi aplicado para identificar os dois atributos mais importantes. O valor absoluto dos coeficientes foi usado para selecionar os dois atributos com maior impacto no modelo.*
- ***Recalcular a Quantidade de Clusters com os Atributos Selecionados:*** *Após selecionar os dois atributos mais relevantes, a análise da quantidade de clusters foi repetida com os métodos do cotovelo e da silhueta, agora considerando apenas esses dois atributos selecionados.*
- ***Comparação entre os Métodos:*** *A quantidade de clusters foi comparada entre os métodos e, caso houvesse divergência, dois scatterplots foram gerados para analisar visualmente as distribuições dos dados com base nos valores de* ***k****.*



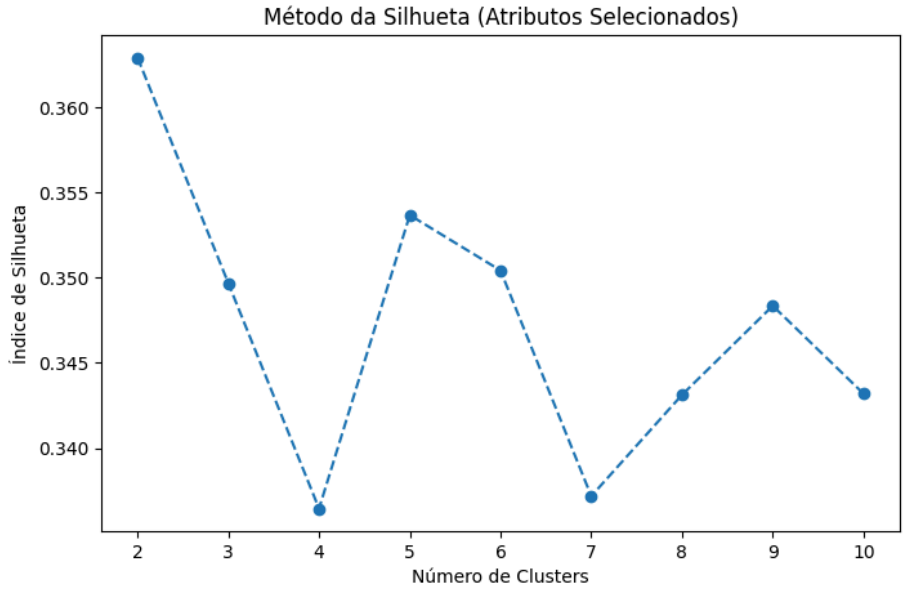
[Figura 7 - Fluxograma do Processo].

- ***Resultado e Análise:***

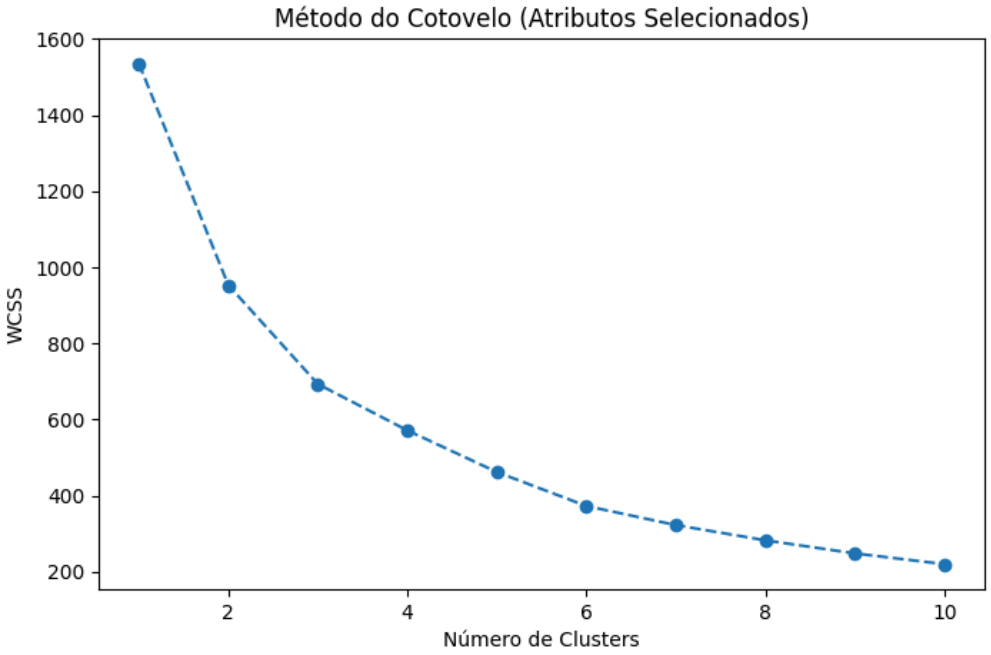
1. ***Atributos Selecionados pelo Lasso****:*
   O Lasso indicou que os dois atributos mais relevantes para a predição foram ***Glucose*** *e* ***BMI****. Esses atributos foram selecionados com base nos coeficientes absolutos mais altos.*
2. ***Método do Cotovelo (Atributos Selecionados)****:*
   O gráfico do método do cotovelo indicou que o número ideal de clusters era k=3. A curva mostrou uma desaceleração significativa da WCSS (Within-Cluster Sum of Squares) entre k=3k e k=4, sugerindo que 3 clusters é um bom valor para o agrupamento.
3. ***Método da Silhueta (Atributos Selecionados)****:*
   O gráfico da silhueta mostrou dois picos: um em k=2 e outro em k=5. No entanto, o valor k=2 teve o maior índice de silhueta, o que sugeriu que dois clusters poderiam ser a melhor escolha para uma separação mais clara dos dados.
4. *O* ***Método do Cotovelo*** *sugeriu k=3 como o número ideal de clusters, devido à redução significativa da WCSS após esse ponto. Isso indicou que o agrupamento em 3 clusters oferece um bom equilíbrio entre a coesão interna e a separação externa dos clusters.*
5. *O* ***Método da Silhueta****, por outro lado, favoreceu k=2, sugerindo que uma segmentação mais simples em dois grupos explicaria melhor a estrutura dos dados. O maior índice de silhueta foi observado com dois clusters, indicando que essa divisão resultou em clusters mais bem separados.*



[Figura 8 - Gráfico da Importância dos Atributos segundo Lasso].



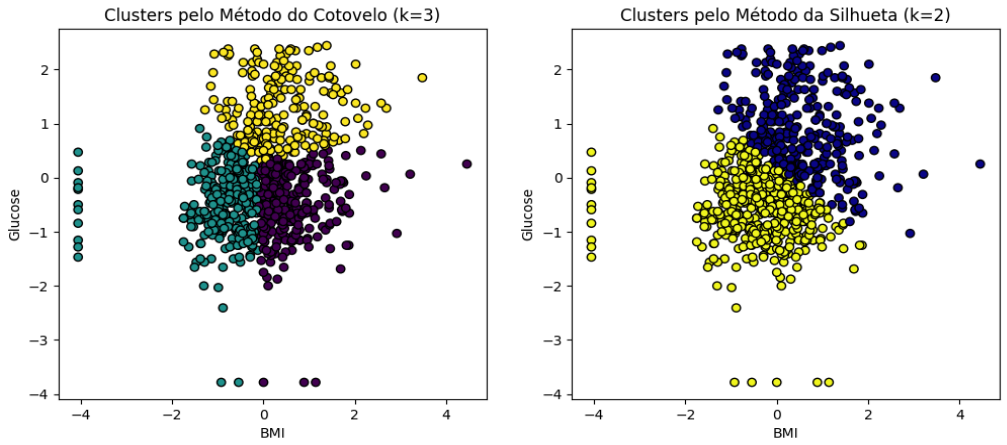
[Figura 9 - Método da Silhueta].



[Figura 10 - Método do Cotovelo].

- ***Comparação Visual - Scatterplots:*** *Para comparar visualmente os resultados dos dois métodos, scatterplots foram gerados para k=3 (Método do Cotovelo) e k=2 (Método da Silhueta).*

1. ***Gráfico do Método do Cotovelo (k = 3)****: O gráfico mostrou três clusters distintos, representados por cores diferentes. Houve alguma sobreposição entre os clusters, mas a separação foi razoavelmente clara, especialmente ao longo dos eixos* ***BMI*** *e* ***Glucose****.*
2. ***Gráfico do Método da Silhueta (k = 2)****: O gráfico mostrou uma separação mais clara entre dois grupos, com uma divisão nítida ao longo do eixo* ***BMI****, onde um cluster representava indivíduos com níveis mais baixos de glicose e o outro com níveis mais altos.*



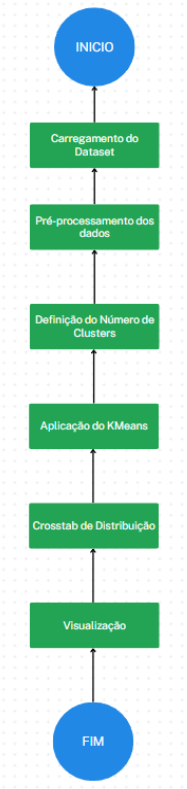
[Figura 11 - Gráfico mostrando número de Clusters por cada Método].

- *O* ***Método do Cotovelo*** *indicou k=3 como o número ideal de clusters, sugerindo uma segmentação mais detalhada, embora com alguma sobreposição entre os grupos.*
- *O* ***Método da Silhueta*** *sugeriu k=2, fornecendo uma divisão mais simples, mas com uma separação mais clara entre os grupos.*
- *A principal conclusão foi que, embora ambos os métodos sugerissem valores diferentes para* ***k****, a análise visual dos scatterplots confirmou que a escolha de k=3 poderia ser mais adequada para capturar a complexidade dos dados, enquanto k=2 forneceu uma separação mais simples, mas bem definida.*

Questão 4 - *Realizar uma análise da distribuição dos clusters obtidos pelo algoritmo KMeans em relação à variável alvo Outcome do dataset. Utilizar o número de clusters determinado pelo índice de silhueta e, a partir disso, criar um crosstab para verificar como os clusters se distribuem entre as classes do Outcome.*

Etapas do Fluxograma do Processo:

- ***Carregamento do Dataset*** *O primeiro passo é carregar o dataset a partir do arquivo* `diabetes.csv`*. Esse dataset contém os dados necessários para a análise, incluindo as variáveis preditoras e a variável alvo "Outcome", que indica se o paciente tem diabetes (1) ou não (0).*
- ***Pré-processamento de Dados*** *Em seguida, as variáveis independentes (features) e a variável dependente (a coluna "Outcome") são separadas. As features são normalizadas utilizando o* ***StandardScaler****, para garantir que todas as variáveis estejam na mesma escala e possam ser processadas de forma mais eficiente no modelo de clustering.*
- ***Definição do Número de Clusters*** *O número ideal de clusters é determinado utilizando o* ***índice de silhueta****, que já foi calculado anteriormente e resultou em 2 clusters como o valor ótimo. Esse índice avalia a qualidade da separação entre os clusters, e o valor obtido indica que 2 clusters são suficientes para capturar as características dos dados.*
- ***Execução do Algoritmo K-Means*** *Com o número de clusters definido, o algoritmo* ***K-Means*** *é aplicado aos dados normalizados. O K-Means tenta agrupar os dados em clusters da forma mais homogênea possível, atribuindo cada observação a um dos dois clusters definidos.*
- ***Geração do Crosstab*** *Após a execução do K-Means, a distribuição dos dados em relação à variável alvo "Outcome" é analisada por meio de uma tabela de contingência (****crosstab****). Essa tabela mostra como os dados de cada cluster estão divididos entre as classes da variável "Outcome", ou seja, quantos indivíduos em cada cluster têm ou não diabetes.*
- ***Visualização do Crosstab*** *Por fim, a distribuição dos clusters em relação ao "Outcome" é apresentada visualmente por meio de um gráfico de barras empilhadas. Esse gráfico facilita a compreensão da distribuição dos dados em cada cluster, mostrando claramente a proporção de pacientes com e sem diabetes dentro de cada grupo formado pelo K-Means.*



[Figura 12 - Fluxograma do Processo].

Tabela Crosstab:

Cluster	Outcome 0	Outcome 1
0	126	144
1	374	124

[Figura 12 - Tabela Crosstab].

A tabela crosstab acima apresenta o número de observações de cada cluster que pertencem às classes "Outcome 0" (não diabéticos) e "Outcome 1" (diabéticos). O número total de observações no dataset é 1000, e a distribuição dos clusters é a seguinte:

- ***Cluster 0*** *contém* ***126*** *indivíduos sem diabetes (Outcome 0) e* ***144*** *indivíduos com diabetes (Outcome 1).*
- ***Cluster 1*** *contém* ***374*** *indivíduos sem diabetes (Outcome 0) e* ***124*** *indivíduos com diabetes (Outcome 1).*

- ***Resultado e Análise:***

Distribuição de Pacientes entre os Clusters: *A crosstab mostra como os pacientes se distribuem entre os dois clusters formados pelo K-Means e como essas distribuições se relacionam com a classe de "Outcome" (diabético ou não). A análise da crosstab pode ser descrita da seguinte forma:*

Cluster 0:

- *Este cluster apresenta uma* ***distribuição equilibrada*** *entre pessoas com e sem diabetes, com* ***144*** *pessoas com diabetes (Outcome 1) e* ***126*** *pessoas sem diabetes (Outcome 0).*
- *Isso sugere que o cluster 0 pode representar um grupo intermediário, possivelmente de pessoas com risco elevado para diabetes, visto que contém uma quantidade considerável de ambos os tipos de resultados (diabéticos e não diabéticos).*

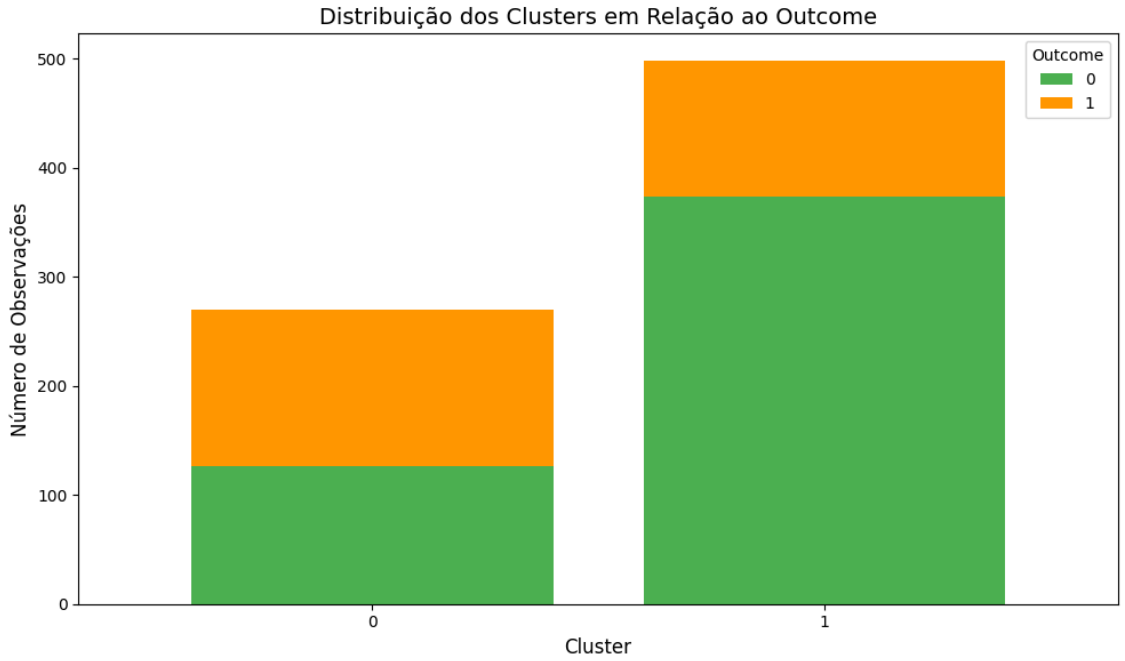
Cluster 1:

- *O* ***Cluster 1****, por outro lado, tem uma* ***maior concentração de pessoas sem diabetes****, com* ***374*** *indivíduos com Outcome 0 e* ***124*** *indivíduos com Outcome 1.*
- *Este cluster parece ser composto principalmente por indivíduos sem diabetes, com uma pequena proporção de pessoas que têm diabetes, o que pode indicar que o modelo tem uma boa capacidade de segmentação e que o Cluster 1 é mais representativo de indivíduos com baixo risco para a doença.*

- ***Interpretação:***

1. *Cluster 0 parece refletir um grupo mais heterogêneo, enquanto o Cluster 1 é mais homogêneo em relação à classe de Outcome 0, ou seja, pessoas sem diabetes.*
2. *O KMeans foi capaz de dividir os pacientes de forma razoável entre os dois clusters, mas, como vimos na análise dos gráficos anteriores (como o Método da Silhueta), o agrupamento pode ainda ser relativamente misturado.*

- ***Potenciais Ações:*** *Com base nesta análise, poderíamos investigar mais a fundo os atributos que levaram a essa divisão, como o Índice de Massa Corporal (BMI) e níveis de glicose, que poderiam ajudar a refinar os clusters e fazer com que as separações entre as classes de Outcome fossem mais claras.*



[Figura 13 - Gráfico da Distribuição dos Clusters em Relação ao Outcome]

- *O número ideal de clusters (2) sugerido pelo índice de silhueta parece fazer sentido quando olhamos a **crosstab**, pois ele dividiu bem os dados entre pessoas com e sem diabetes, mas ainda há uma certa sobreposição, principalmente no **Cluster 0**. Para um estudo mais refinado, poderia ser interessante explorar valores de **k** diferentes e testar outras abordagens, como a utilização de mais variáveis para o agrupamento.*

4. Conclusões

Os resultados esperados eram:

- *Analisar a relevância dos atributos para o modelo de classificação de diabetes utilizando a técnica de Lasso.*
- *Determinar o número ideal de clusters para a segmentação dos dados utilizando os métodos do Cotovelo e da Silhueta.*
- *Visualizar a distribuição dos clusters em relação à variável alvo "Outcome", que indica a presença ou ausência de diabetes, utilizando a técnica de K-Means.*
- *Analisar a distribuição dos clusters com base nas classes de "Outcome" e sua relação com os resultados obtidos.*

Os resultados obtidos indicam que:

Atributos Importantes:

- *O uso do Lasso indicou que os dois atributos mais importantes para o modelo de predição foram "Glucose" e "BMI" (Índice de Massa Corporal). Esses dois atributos têm um impacto considerável na detecção da presença ou ausência de diabetes, conforme os coeficientes absolutos obtidos na análise do Lasso. O gráfico de importância também destacou claramente essa relação.*

Determinação do Número de Clusters:

- *Utilizando os métodos do Cotovelo e da Silhueta, foi possível identificar o número ideal de clusters para a segmentação dos dados. O método da Silhueta sugeriu que o número ideal de clusters seria 2, enquanto o Método do Cotovelo indicou a possibilidade de 3 clusters. A análise mostrou que, embora a segmentação em 2 clusters (indicado pela Silhueta) fosse mais clara e apresentasse uma separação mais evidente, o Método do Cotovelo sugeria que um número maior de clusters poderia ser relevante para identificar subgrupos dentro do conjunto de dados. Ambos os resultados foram válidos, mas a escolha de 2 clusters, com base no Índice de Silhueta, parece ser mais robusta neste contexto.*

Distribuição dos Clusters com relação ao "Outcome":

- *O K-Means com 2 clusters (segundo o método da silhueta) foi capaz de segmentar bem os dados, mas com certa sobreposição. A Crosstab gerada entre os clusters e o Outcome revelou que o Cluster 0 continha uma distribuição equilibrada entre pessoas com e sem diabetes, enquanto o Cluster 1 apresentou uma maior concentração de indivíduos sem diabetes. Isso indica que o modelo de agrupamento separou parcialmente os pacientes com diabetes dos sem diabetes, mas ainda há uma margem para melhoria na separação de grupos com maior clareza.*
- *A análise também destacou que o **Cluster 0** pode representar um grupo mais heterogêneo, enquanto o **Cluster 1** parece ser mais homogêneo.*

Análise dos Resultados

Embora os resultados estejam de acordo com as expectativas, a segregação entre os clusters não foi tão nítida quanto se esperava. Isso se deve a uma combinação de fatores:

- *O K-Means pode não ser o método de clustering mais eficiente para dados com muitas sobreposições e variáveis de alta correlação, como ocorre em dados de saúde.*
- *Os clusters encontrados (especialmente o Cluster 0) são relativamente misturados, sugerindo que a diferenciação entre pessoas com e sem diabetes pode ser mais complexa do que apenas a segmentação em dois ou três grupos. Uma análise mais profunda da distribuição das variáveis pode fornecer insights adicionais para uma segmentação mais precisa.*

No entanto, de maneira geral, os resultados foram satisfatórios, pois a análise forneceu insights válidos sobre a importância dos atributos, a quantidade de clusters e a segmentação dos dados, permitindo uma compreensão mais detalhada do comportamento dos pacientes em relação à diabetes.

5. Próximos passos

Refinamento da Modelagem de Clusters:

- *Aprimorar a técnica de clustering utilizando outros métodos além do K-Means, como o DBSCAN (para clusters de forma arbitrária) ou Hierarchical Clustering. Esses métodos podem oferecer uma separação mais nítida entre os clusters, especialmente quando há grandes sobreposições nos dados.*
- *Testar outros valores de k para o K-Means, ajustando melhor o número de clusters e analisando a distribuição dos dados em diferentes cenários.*
- *Explorar técnicas de redução de dimensionalidade, como PCA (Principal Component Analysis) ou t-SNE, para entender como a segmentação dos clusters se comporta em um espaço de menor dimensão.*

***Aprimoramento da Seleção de Atributos**:*

- *Realizar uma **seleção de atributos mais robusta**, possivelmente utilizando outras técnicas além do **Lasso**, como **Random Forests** ou **Gradient Boosting Machines** para identificar os atributos mais relevantes.*
- *Investigar a **correlação entre os atributos**, para verificar se algumas variáveis podem ser combinadas ou descartadas, melhorando a qualidade do modelo.*

***Avaliação e Validação do Modelo**:*

- *Validar o modelo de clustering utilizando uma base de dados adicional ou usando a **validação cruzada** para verificar a generalização dos resultados.*
- *Implementar métricas adicionais de avaliação para a segmentação, como a **Índice de Davies-Bouldin**, que mede a separação entre os clusters, ou outras métricas de qualidade de clustering.*

***Integração com Outras Técnicas de Machine Learning**:*

- *Explorar a utilização do **KMeans** como uma etapa de pré-processamento para agrupamento de dados, para então aplicar modelos de **classificação supervisionada** (como **Random Forests** ou **SVM**) que possam usar os clusters como novas variáveis de entrada. Isso pode melhorar a precisão do modelo preditivo para a presença ou ausência de diabetes.*
- *Analisar as **previsões do modelo** com base em outras técnicas de machine learning para comparar a eficácia do KMeans na segmentação dos dados.*

***Aplicação Prática**:*

- *Considerando que a segmentação dos dados pode ter implicações para a **saúde pública**, os **próximos passos** podem incluir a utilização do modelo para prever o risco de diabetes em novos pacientes com base nas variáveis analisadas, o que pode ser útil em clínicas, hospitais ou centros de saúde.*
- ***Testar o modelo** em diferentes populações, variando os dados de entrada, para verificar a generalização do modelo e ajustar as previsões de acordo com diferentes perfis de pacientes.*

*Em suma, os próximos passos buscam não apenas aprimorar a análise dos dados com técnicas mais avançadas, mas também aplicar esses modelos de maneira prática para obter **resultados mais precisos** e **utilizáveis** em contextos reais de saúde.*