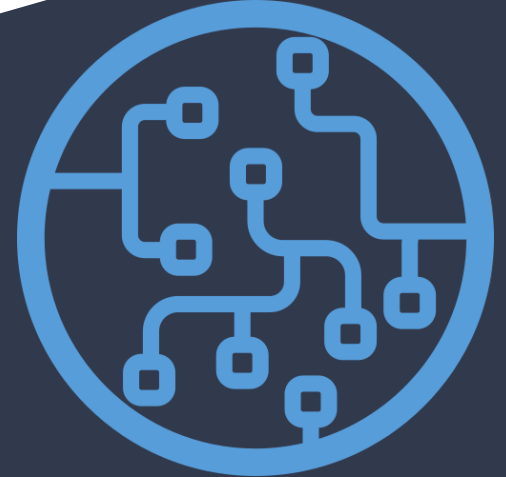


# Modelo Preditivo para Diagnóstico de Doença Hepática

Análise, Modelagem e Comparação de Algoritmos



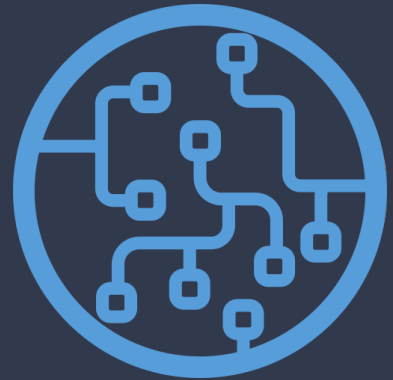
# Problema de Negócio:

- A doença hepática e o problema na área da saúde pública e privada
- Dificuldade na predição
- Detecção tardia
- Custos dos atuais métodos
- Métodos invasivos e demorados



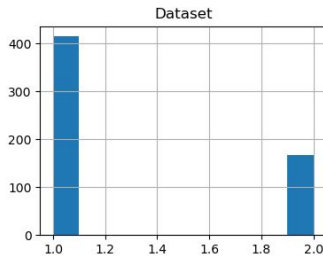
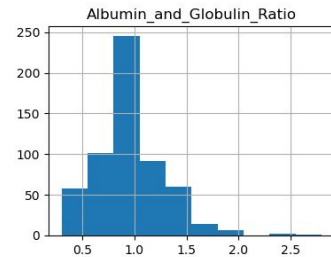
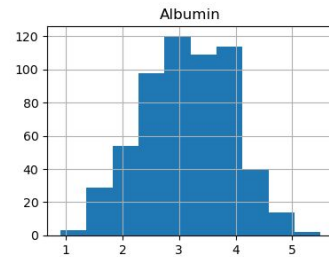
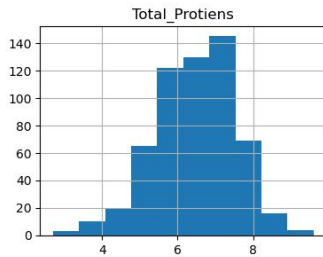
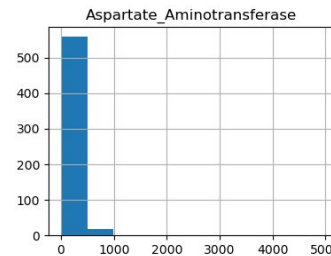
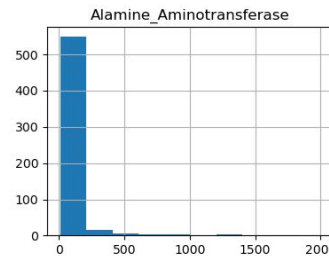
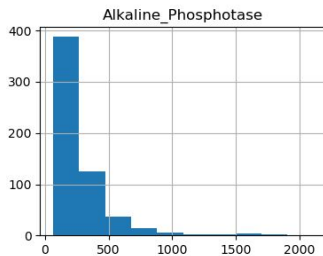
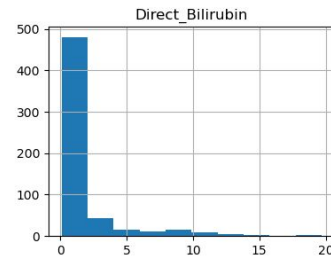
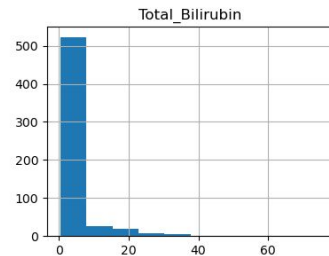
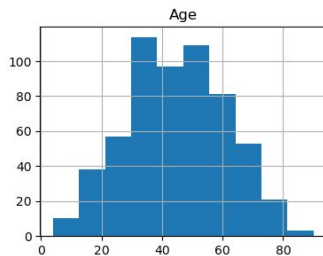
# Objetivo:

- Desenvolver e avaliar diferentes modelos de Machine Learning
- Compreender influência das Variáveis Predictoras
- Escolher o melhor modelo preditivo
  - Identificar pacientes com alta probabilidade de possuir uma doença hepática
  - Baseado em variáveis fisiológicas.
- Testar com novos dados



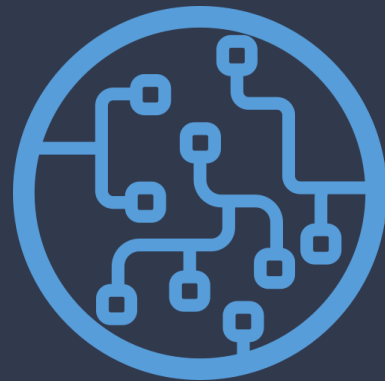
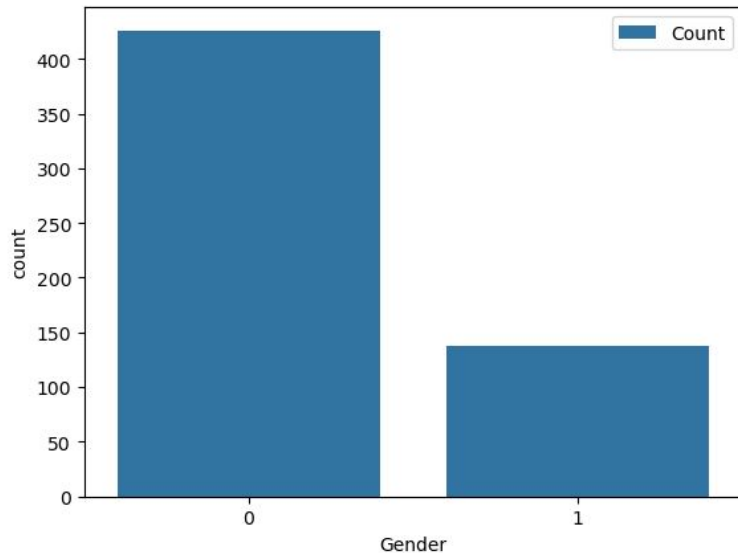
# Análise Exploratória

- Outliers a Tratar
  - Total\_Bilirubin
  - Direct\_Bilirubin
  - Alkaline\_Phosphotase
  - Alamine\_Aminotransferase
  - Aspartate\_Aminotransferase
- Soluções tomadas
  - Remoção de registros a partir de um n° de corte como sugerido pela equipe



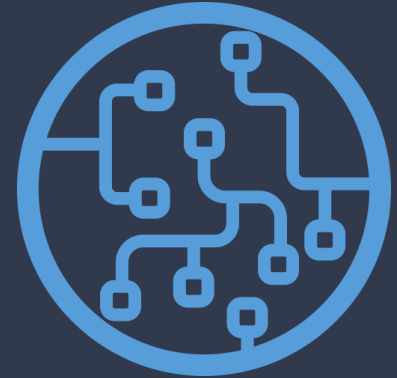
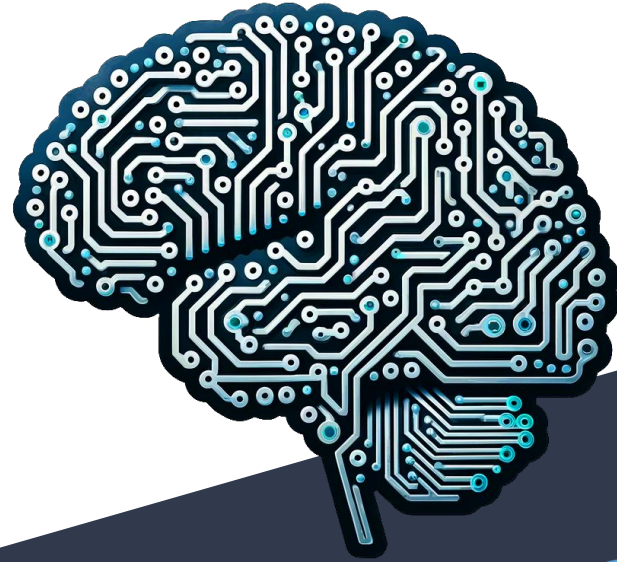
# Análise Exploratória

- Desbalanceamento de Classes
  - Número de pacientes do sexo masculino: 426 (0)
  - Número de pacientes do sexo feminino: 138 (1)
- Soluções tomadas
  - Oversampling com SMOTE



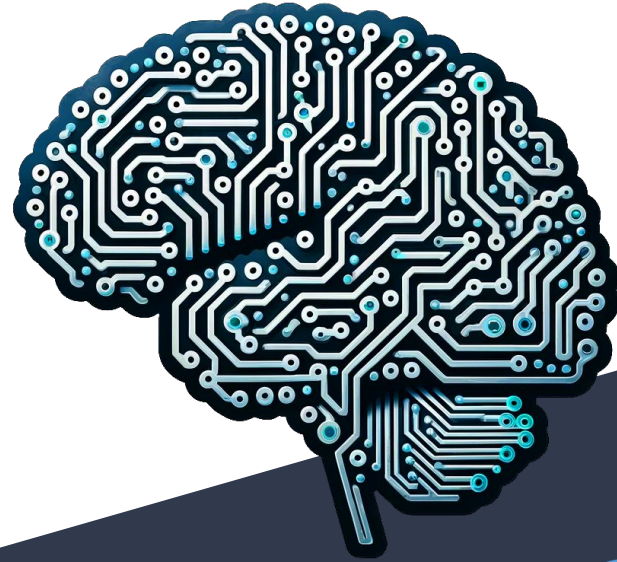
# Machine Learning

- Divisão para Treino e Teste
  - 75% para Treino
  - 25% para Teste
- Padronização dos Dados



# Machine Learning

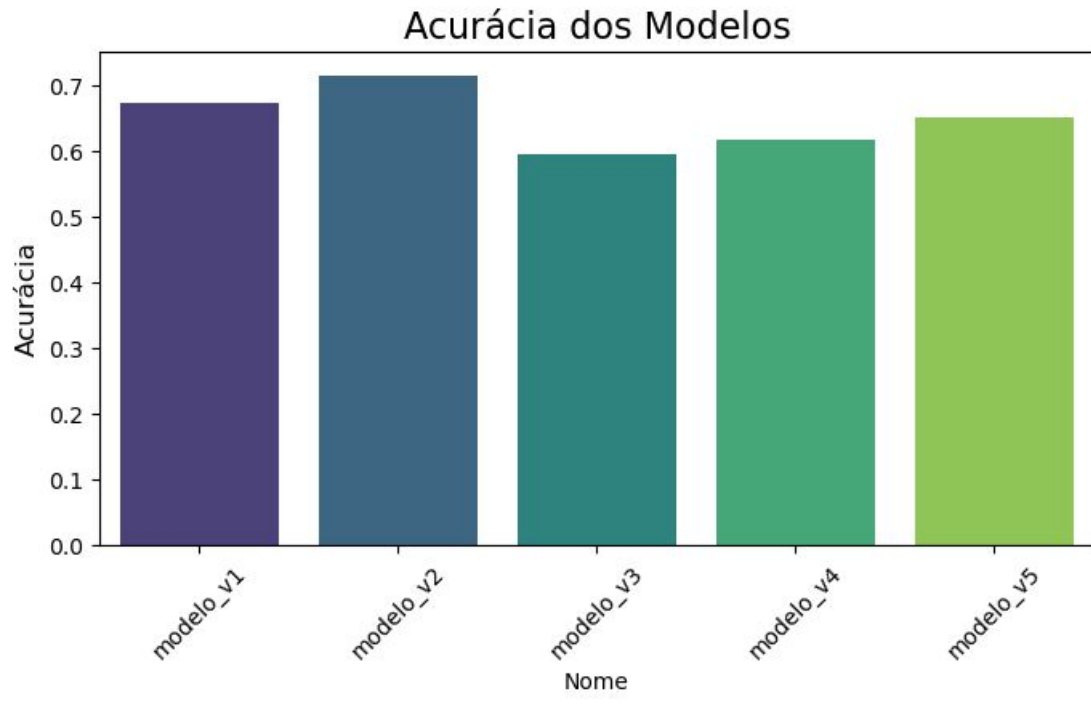
- Modelo V1
  - Regressão Logística
- Modelo V2
  - Random Forest
- Modelo V3
  - KNN
- Modelo V4
  - Decision Tree
- Modelo V5
  - SVM



# Machine Learning

o Random Forest se destacou como o mais eficiente, com um bom equilíbrio entre complexidade computacional e desempenho preditivo.

A Acurácia foi de 0.71, sugerindo que o modelo tem uma taxa de sucesso razoável na classificação correta dos casos.





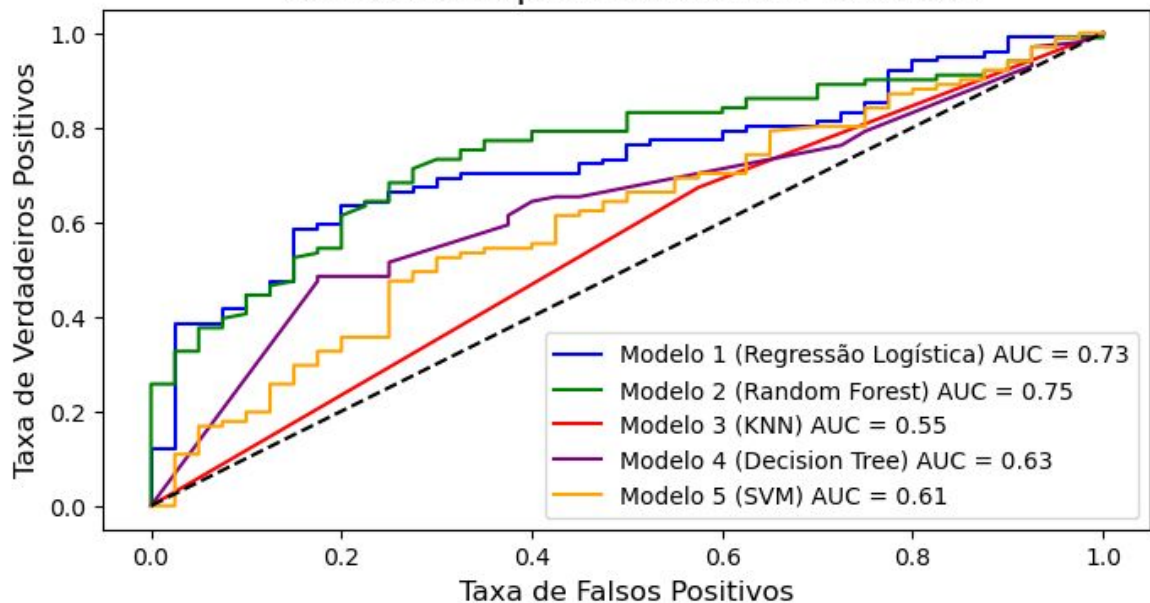
# Machine Learning

Cada curva mostra a relação entre Falsos Positivos e Verdadeiros Positivos

O valor AUC indica a qualidade do modelo

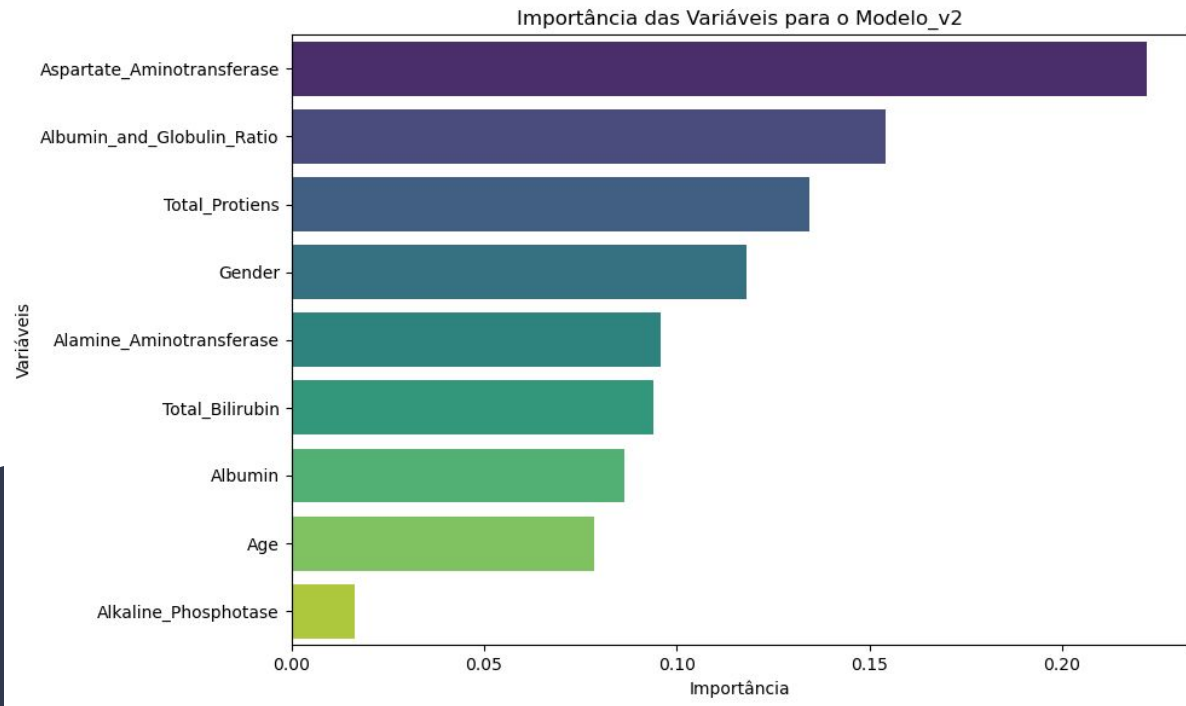
O modelo Random Forest obteve AUC de 0.75, indicando uma boa capacidade de discriminação entre pacientes com e sem risco de doença hepática.

Curvas ROC para Diferentes Modelos



# Feature Importance

As métricas de importância das variáveis ajudam a identificar quais fatores fisiológicos influenciam mais a classificação.



# Testando o Modelo com Novos Dados

Um exemplo de um novo paciente foi classificado utilizando o modelo de Random Forest.

A probabilidade de desenvolver a doença hepática foi predita com base nas características fornecidas, com um resultado indicando a presença de risco para doença hepática.

## Novas Previsões com o Random Forest

```
# Novo paciente
novo_paciente = [72, 1, 0.8, 186, 15, 29, 7.1, 3.4, 0.97]
```

```
# Conv. em array
arr_paciente = np.array(novo_paciente)
```

```
# Padronização
arr_paciente = (arr_paciente - treino_mean) / treino_std
```

```
arr_paciente
```

Age	1.756525
Gender	1.942036
Total_Bilirubin	-0.336404
Alkaline_Phosphotase	-0.366761
Alamine_Aminotransferase	-0.340191
Aspartate_Aminotransferase	-0.344365
Total_Protiens	0.576059
Albumin	0.200208
Albumin_and_Globulin_Ratio	-0.116449
dtype:	float64

```
# Converter para array NumPy e redimensionar corretamente
arr_paciente = np.array(arr_paciente).reshape(1, -1)
```

```
# Fazer a previsão
pred_novo_paciente = modelo_v2.predict(arr_paciente)
```

```
# Resultado
if pred_novo_paciente == 1:
    print('Este paciente pode apresentar doença hepática!')
else:
    print('Não há alta probabilidade de apresentar doença hepática!')
```

Este paciente pode apresentar doença hepática!

Aberto a Dúvidas e  
Sugestões. Muito Obrigado!

