



AVALIAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE MORTALIDADE NEONATAL UTILIZANDO DADOS DO DATASUS.

Aluno: Caio Augusto de Souza Mota
Orientador Prof. Me. Everton Josué da Silva

- Taxa de mortalidade infantil.
- Mortalidade Neonatal 60% da TMI.
- Experiência do profissional.

Objetivo Geral:

- O presente trabalho tem como objetivo testar e analisar dois tipos diferentes de algoritmos de aprendizagem de máquina, utilizando a base de dados do SIM e SINASC do Brasil inteiro para predição de morte neonatal.

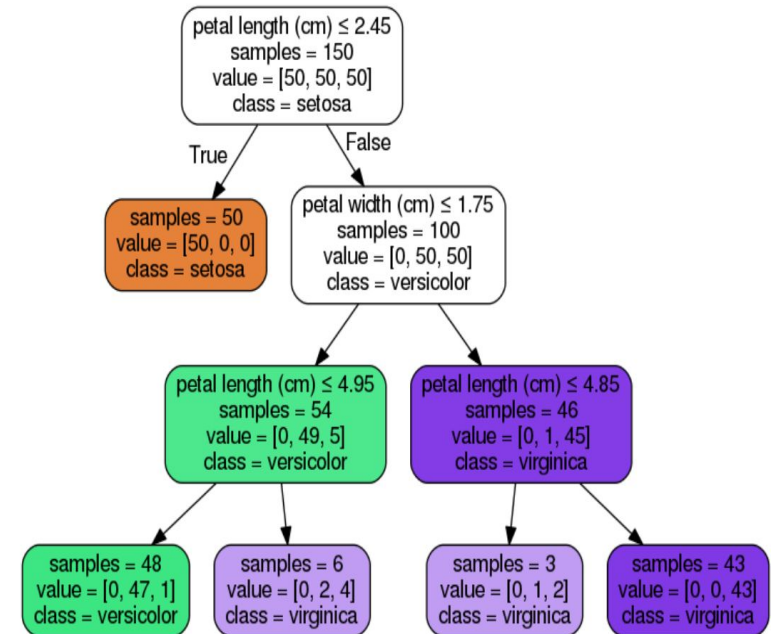
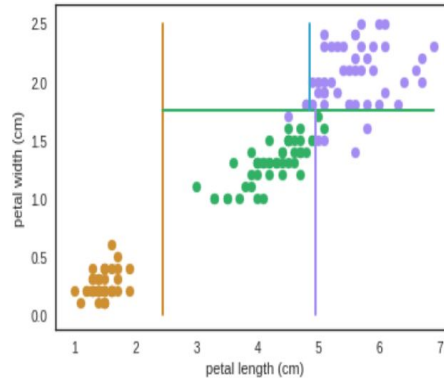
Objetivos Específicos:

- Desenvolver/Alterar cada um dos algoritmos propostos (Árvore de Decisão e Regressão Logística);
- Realizar Análise exploratória dos dados;
- Realizar tratamento e particionamento dos dados para serem aplicados nos treinos e testes dos modelos;
- Treinar e Testar os algoritmos para gerar os modelos;
- Analisar os resultados.

- Aplicação do modelo é inovador.
- Redução da mortalidade neonatal.

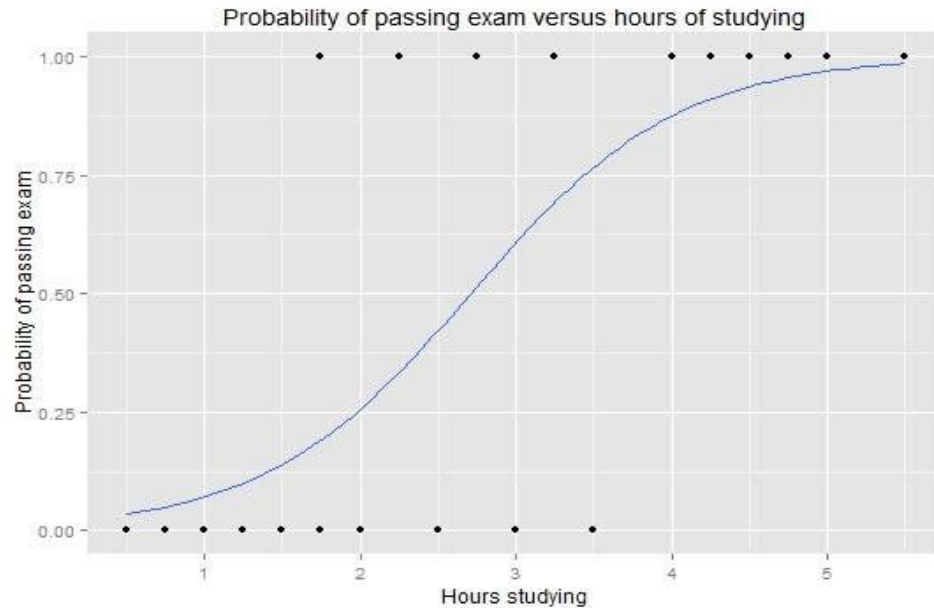
Árvore de decisão:

- Supervisionado.
- Classificação.
- Problema complexo é decomposto em subproblemas.
- WhiteBox.



Regressão Logística:

- Supervisionado.
- Classificação.
- Utiliza a Função Linear.
- Função Logística.



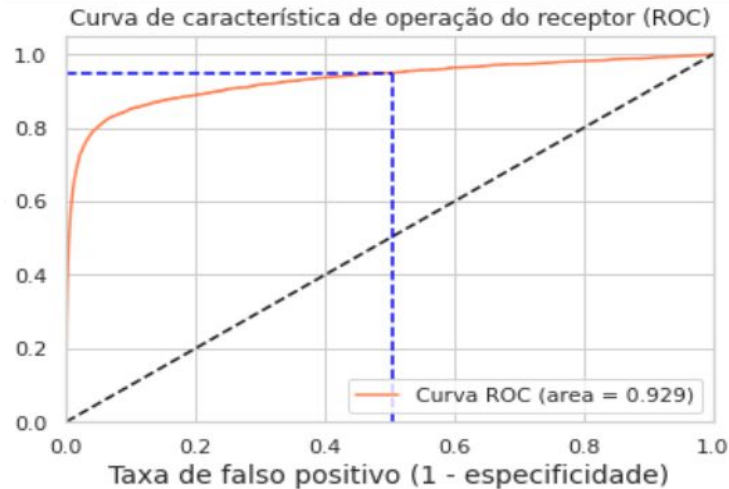
Treinamento e Teste:

- 90% para treino e 10% para os testes.
- K-folds cross-validation.

| Predição 1 | Predição 2 | Predição 3 | Predição 4 | Predição 5 |
|------------|------------|------------|------------|------------|
| Teste | Treino | Treino | Treino | Treino |
| Treino | Teste | Treino | Treino | Treino |
| Treino | Treino | Teste | Treino | Treino |
| Treino | Treino | Treino | Teste | Treino |
| Treino | Treino | Treino | Treino | Teste |

Avaliação dos modelos:

Curva ROC



Matriz de Confusão

| | | Predito | |
|-------------|--------------|---------------------|---------------------|
| | | Negativo (0) | Positivo (1) |
| Classe Real | Negativo (0) | Verdadeiro Negativo | Falso Positivo |
| | Positivo (1) | Falso Negativo | Verdadeiro Positivo |

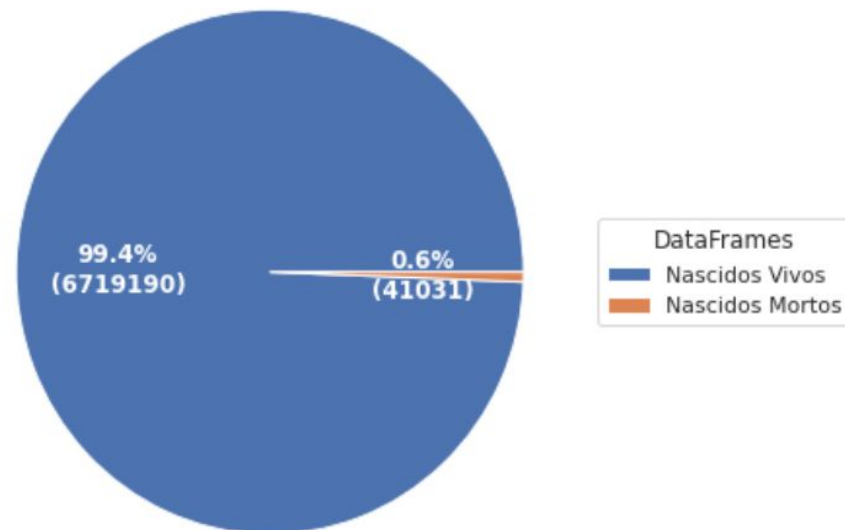
Ferramentas utilizadas:

- Linguagem: Python.
- Plataformas: Jupyter Notebook e Google Colab.

Base de dados:

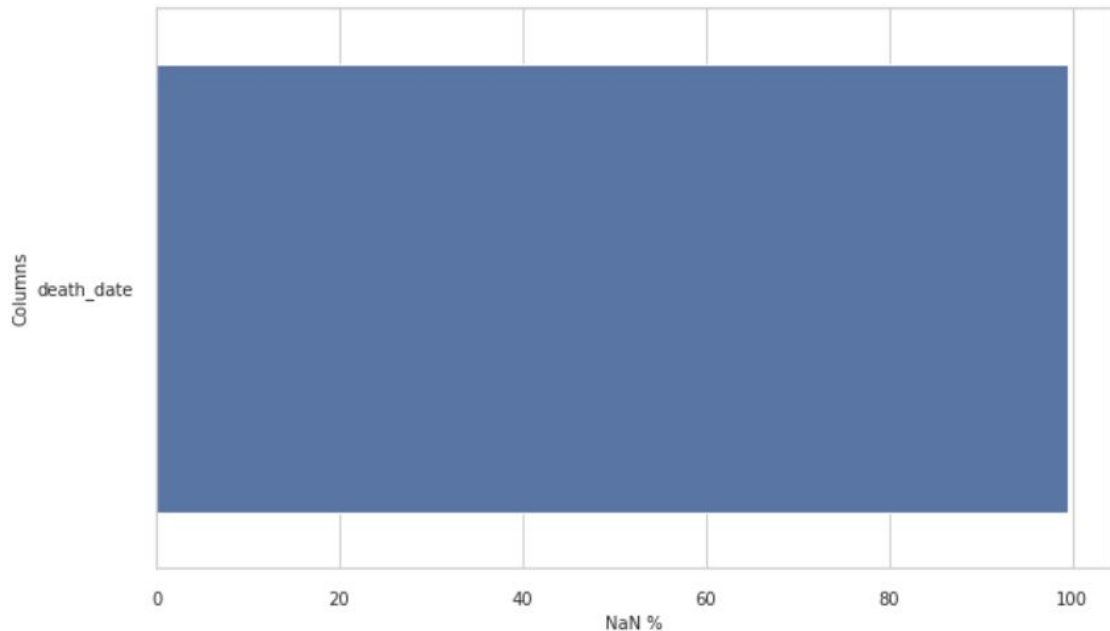
- Sistema de Informação sobre Mortalidade (SIM) e Sistema de Informações sobre Nascidos Vivos (SINASC) do Brasil inteiro, entre os anos 2016 ate 2018.

Porcentagem de amostras de Nascidos Vivos e Nascidos Mortos



Preparação da base de dados:

- Limpeza dos dados.
- Somente uma variável tem valores nulos.
- Variáveis retiradas.



Análise Exploratória:

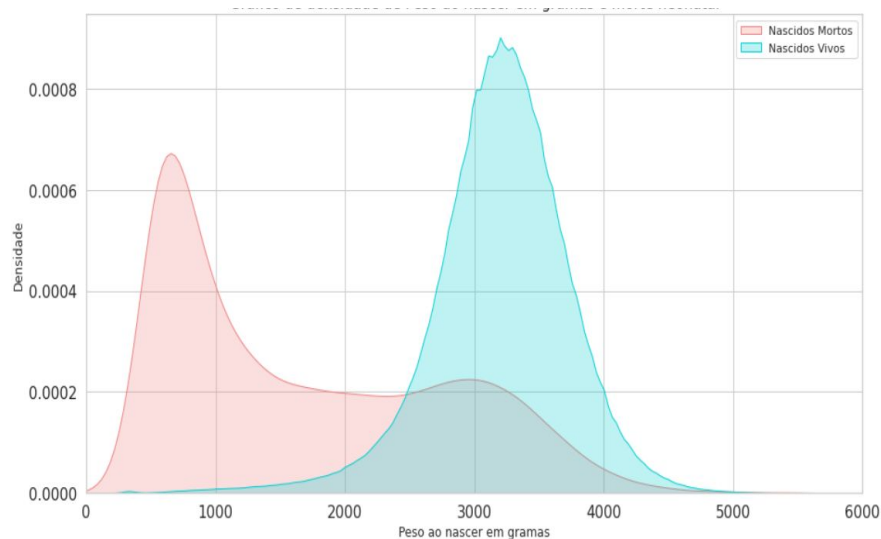


Gráfico de densidade de nascidos vivos e nascidos mortos usando Peso ao Nascer.

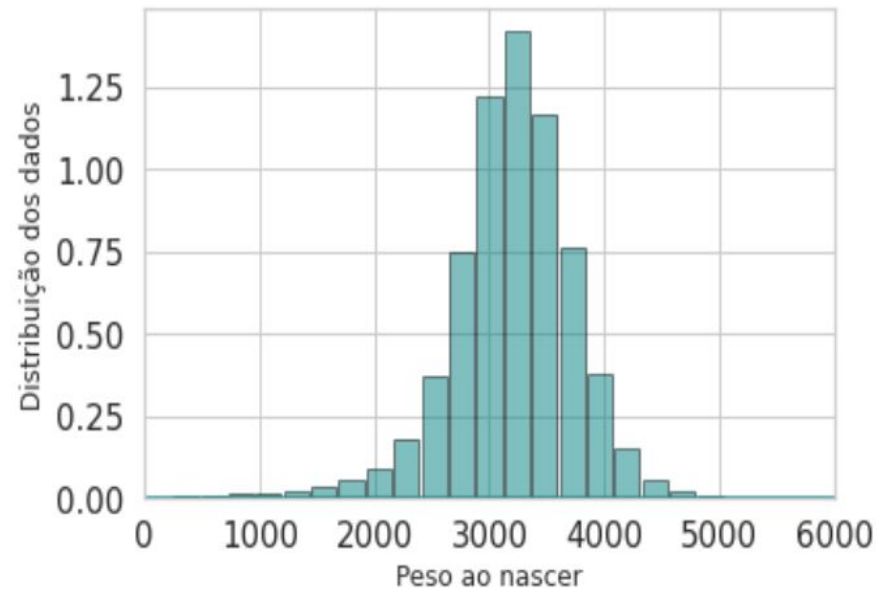
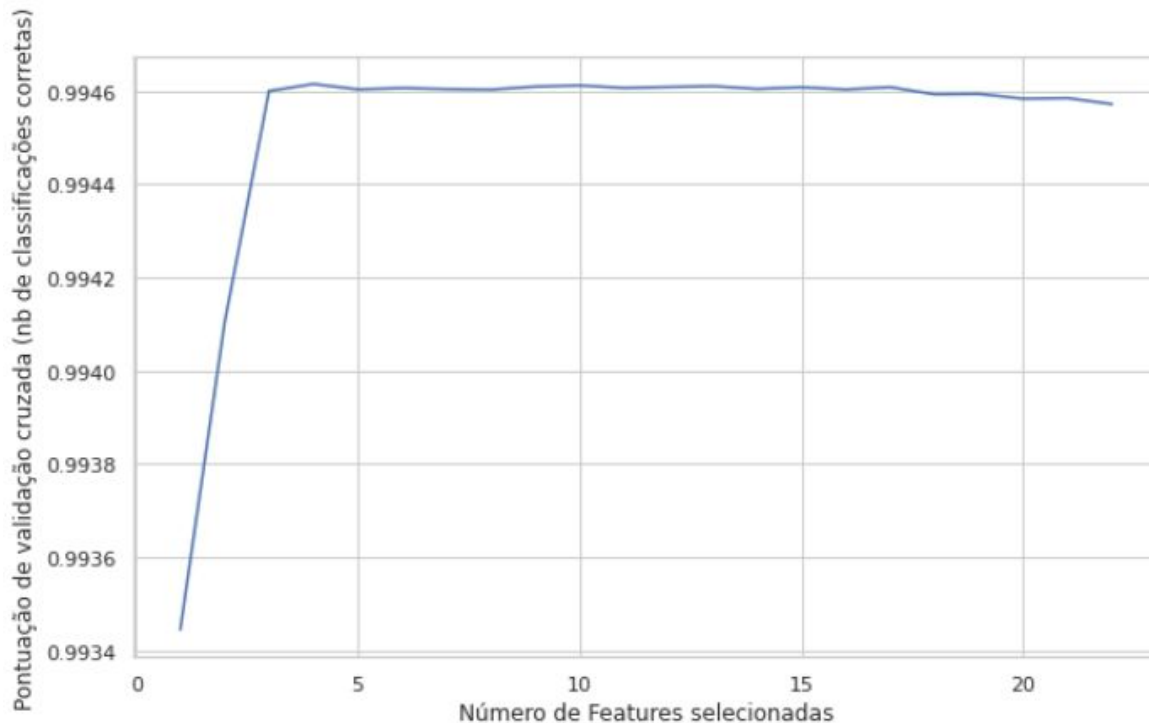


Gráfico Distribuição dos dados de Peso ao Nascer.

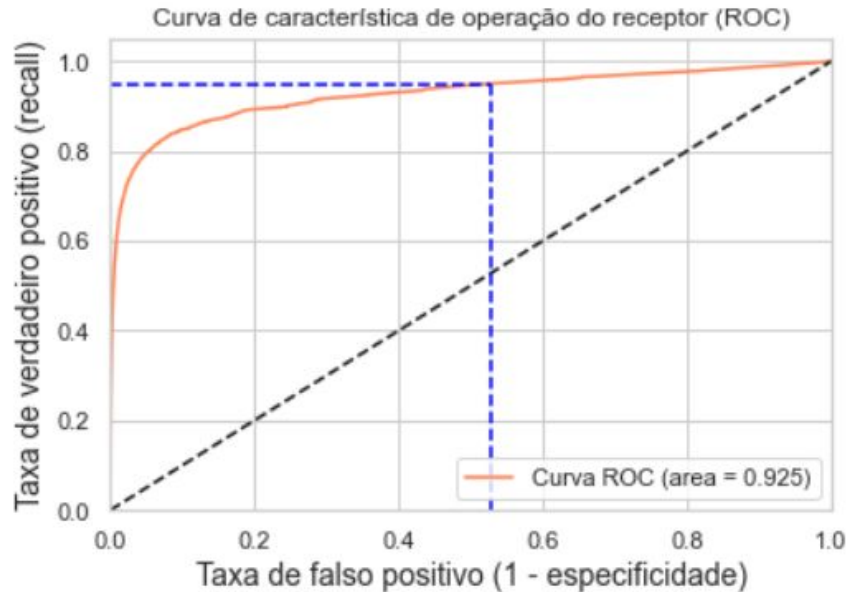
| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|-----------------------------|---------|---------|---------|----------|----------|---------|---------|----------|---------|----------|----------|---------|----------|---------|---------|----------|----------|---------|---------|---------|----------|---------|---------|
| NÚMERO DE PARTOS NORMAIS (TRABALHOS DE PARTO) | Local de nascimento | 1 | -0.0044 | 0.0086 | -0.019 | 0.055 | -0.00035 | -0.025 | -0.01 | -0.087 | -0.08 | -0.033 | -0.06 | -0.023 | -0.0034 | 0.0059 | 0.043 | 0.067 | -0.022 | -0.0017 | -0.0056 | 0.18 | -0.021 | 0.0084 | 0.0084 |
| | Idade da Mãe | -0.0044 | 1 | 0.088 | 0.27 | 0.31 | 0.18 | -0.024 | 0.068 | 0.29 | 0.14 | 0.038 | 0.034 | 0.014 | -0.011 | -0.12 | 0.32 | 0.17 | 0.25 | -0.073 | 0.054 | -0.13 | -0.042 | 0.17 | -0.0054 |
| | Estado civil da mãe | 0.0086 | 0.088 | 1 | -0.008 | 0.077 | 0.053 | -0.0031 | 0.015 | -0.0013 | 0.007 | 0.0046 | 0.004 | 0.011 | -0.0088 | 0.078 | 0.09 | 0.088 | 0.067 | 0.0075 | 0.13 | 0.078 | 0.097 | 0.077 | 0.0093 |
| | | Anos de escolaridade da mãe | -0.019 | 0.27 | -0.008 | 1 | -0.29 | -0.026 | -0.0073 | 0.044 | 0.3 | 0.19 | 0.076 | 0.083 | -0.0074 | -0.0024 | -0.22 | -0.24 | -0.29 | 0.016 | -0.06 | 0.067 | -0.13 | -0.031 | -0.0098 |
| | Número de nascidos vivos | 0.055 | 0.31 | 0.077 | -0.29 | 1 | 0.15 | -0.0021 | 0.0032 | -0.12 | -0.17 | 0.014 | -0.0083 | 0.049 | -0.00041 | 0.093 | 0.84 | 0.75 | 0.39 | 0.0018 | -0.0092 | 0.074 | 0.023 | 0.24 | 0.0029 |
| | Número de perdas fetais | -0.00035 | 0.18 | 0.053 | -0.026 | 0.15 | 1 | -0.035 | 0.016 | 0.029 | 0.0027 | -0.011 | -0.014 | -0.024 | -0.0098 | 0.048 | 0.54 | 0.16 | 0.11 | -0.041 | 0.094 | 0.051 | 0.062 | 0.11 | 0.017 |
| | Semanas gestacionais por intervalos | -0.025 | -0.024 | -0.0031 | -0.0073 | -0.0021 | -0.035 | 1 | -0.25 | -0.032 | 0.19 | 0.25 | 0.25 | 0.53 | 0.045 | 0.0065 | -0.016 | -0.011 | 0.0024 | 0.76 | -0.077 | 0.024 | -0.013 | -0.53 | -0.27 |
| | Tipo de gravidez | -0.01 | 0.068 | 0.015 | 0.044 | 0.0032 | 0.016 | -0.25 | 1 | 0.11 | 0.0025 | -0.065 | -0.063 | -0.28 | -0.008 | -0.026 | 0.0084 | 0.00076 | 0.0041 | -0.25 | 0.11 | -0.05 | -0.0015 | 0.22 | 0.046 |
| | Tipo de parto | -0.087 | 0.29 | -0.0013 | 0.3 | -0.12 | 0.029 | -0.032 | 0.11 | 1 | 0.15 | -6.7e-05 | -0.0029 | 0.0021 | -0.015 | -0.081 | -0.29 | 0.28 | -0.085 | 0.076 | -0.39 | -0.057 | 0.12 | -0.0029 | |
| | Número de consultas de pré-natal por faixas | -0.08 | 0.14 | 0.007 | 0.19 | -0.17 | 0.0027 | 0.19 | 0.0025 | 0.15 | 1 | 0.094 | 0.11 | 0.17 | 0.0067 | -0.077 | -0.14 | -0.17 | -0.0086 | 0.17 | -0.0094 | -0.086 | -0.014 | -0.13 | -0.086 |
| | Pontuação de Apgar de 1 minuto | -0.033 | 0.038 | 0.0046 | 0.076 | 0.014 | -0.011 | 0.25 | -0.065 | -6.7e-05 | 0.094 | 1 | 0.74 | 0.19 | 0.072 | -0.037 | 0.0071 | 0.0046 | 0.013 | 0.23 | -0.044 | 0.014 | -0.0098 | -0.11 | -0.23 |
| | Pontuação de Apgar de 5 minuto | -0.06 | 0.034 | 0.004 | 0.083 | -0.0083 | -0.014 | 0.25 | -0.063 | -0.0029 | 0.11 | 0.74 | 1 | 0.19 | 0.076 | -0.034 | -0.011 | -0.0081 | -0.00076 | 0.23 | -0.031 | 0.016 | -0.0037 | -0.12 | -0.26 |
| | Peso ao nascer em gramas | -0.023 | 0.014 | 0.011 | -0.0074 | 0.049 | -0.024 | 0.53 | -0.28 | 0.0021 | 0.17 | 0.19 | 0.19 | 1 | 0.051 | 0.017 | 0.029 | 0.021 | 0.043 | 0.59 | -0.065 | 0.016 | -0.00041 | -0.3 | -0.22 |
| | Presença de malformação congênita | -0.0034 | -0.011 | -0.0088 | -0.0024 | -0.00041 | -0.0098 | 0.045 | -0.008 | -0.015 | 0.0067 | 0.072 | 0.076 | 0.051 | 1 | 0.002 | -0.0033 | -0.00041 | -0.0032 | 0.045 | -0.02 | 0.00036 | -0.0083 | -0.055 | -0.13 |
| | Raça / cor da pele da mãe | 0.0059 | -0.12 | 0.078 | -0.22 | 0.093 | 0.048 | 0.0065 | -0.026 | -0.15 | -0.077 | -0.037 | -0.034 | 0.017 | 0.002 | 1 | 0.092 | 0.2 | 0.047 | 0.074 | 0.37 | 0.42 | 0.36 | 0.15 | 0.015 |
| | Número de gestações anteriores | 0.043 | 0.32 | 0.09 | -0.24 | 0.84 | 0.54 | -0.016 | 0.0084 | -0.081 | -0.14 | 0.0071 | -0.011 | 0.029 | -0.0033 | 0.092 | 1 | 0.74 | 0.39 | -0.017 | 0.011 | 0.072 | 0.031 | 0.25 | 0.0093 |
| | Número de partos normais (trabalhos de parto) | 0.067 | 0.17 | 0.088 | -0.29 | 0.75 | 0.16 | -0.011 | 0.00076 | -0.29 | -0.17 | 0.0046 | -0.0081 | 0.021 | -0.00041 | 0.2 | 0.74 | 1 | -0.088 | 0.024 | 0.13 | 0.27 | 0.15 | 0.16 | 0.011 |
| | Número de partos cesáreos (partos) | -0.022 | 0.25 | 0.067 | 0.016 | 0.39 | 0.11 | 0.0024 | 0.0041 | 0.28 | -0.0086 | 0.013 | -0.00076 | 0.043 | -0.0032 | 0.047 | 0.39 | -0.088 | 1 | -0.021 | 0.14 | -0.013 | 0.091 | 0.33 | 0.0016 |
| | Semana de gestação (por intervalos) | -0.0017 | -0.073 | 0.0075 | -0.06 | 0.0018 | -0.041 | 0.76 | -0.25 | -0.085 | 0.17 | 0.23 | 0.23 | 0.59 | 0.045 | 0.074 | -0.017 | 0.024 | -0.021 | 1 | -0.039 | 0.1 | 0.042 | -0.47 | -0.3 |
| | Tipo de apresentação de recém-nascido | -0.0056 | 0.054 | 0.13 | 0.067 | -0.0092 | 0.094 | -0.077 | 0.11 | 0.076 | -0.0094 | 0.044 | -0.031 | -0.065 | -0.02 | 0.37 | 0.011 | 0.13 | 0.14 | -0.039 | 1 | 0.53 | 0.55 | 0.39 | 0.053 |
| Assistência ao parto | 0.18 | -0.13 | 0.078 | -0.13 | 0.074 | 0.051 | 0.024 | -0.05 | -0.39 | -0.086 | 0.014 | 0.016 | 0.016 | 0.00036 | 0.42 | 0.072 | 0.27 | -0.013 | 0.1 | 0.53 | 1 | 0.47 | 0.15 | 0.0074 | |
| Principal função do trabalhador | -0.021 | -0.042 | 0.097 | -0.031 | 0.023 | 0.062 | -0.013 | -0.0015 | -0.057 | -0.014 | -0.0098 | -0.037 | -0.00041 | -0.0083 | 0.36 | 0.031 | 0.15 | 0.091 | 0.042 | 0.55 | 0.47 | 1 | 0.22 | 0.019 | |
| Classificação do grupo Robson | 0.0084 | 0.17 | 0.077 | -0.0098 | 0.24 | 0.11 | -0.53 | 0.22 | 0.12 | -0.13 | -0.11 | -0.12 | -0.3 | -0.055 | 0.15 | 0.25 | 0.16 | 0.33 | -0.47 | 0.39 | 0.15 | 0.22 | 1 | 0.11 | |
| Morte antes de 28 dias (rótulo) | 0.0084 | -0.0054 | 0.0093 | -0.014 | 0.0029 | 0.017 | -0.27 | 0.046 | -0.0029 | -0.086 | -0.23 | -0.26 | -0.22 | -0.13 | 0.015 | 0.0093 | 0.011 | 0.0016 | -0.3 | 0.053 | 0.0074 | 0.019 | 0.11 | 1 | |
| | Local de nascimento | | | | | | | | | | | | | | | | | | | | | | | | |
| | Idade da Mãe | | | | | | | | | | | | | | | | | | | | | | | | |
| | Estado civil da mãe | | | | | | | | | | | | | | | | | | | | | | | | |
| | Anos de escolaridade da mãe | | | | | | | | | | | | | | | | | | | | | | | | |
| | Número de nascidos vivos | | | | | | | | | | | | | | | | | | | | | | | | |
| | Número de perdas fetais | | | | | | | | | | | | | | | | | | | | | | | | |
| | Semanas gestacionais por intervalos | | | | | | | | | | | | | | | | | | | | | | | | |
| | Tipo de gravidez | | | | | | | | | | | | | | | | | | | | | | | | |
| | Tipo de parto | | | | | | | | | | | | | | | | | | | | | | | | |
| | Número de consultas de pré-natal por faixas | | | | | | | | | | | | | | | | | | | | | | | | |
| | Pontuação de Apgar de 1 minuto | | | | | | | | | | | | | | | | | | | | | | | | |
| | Pontuação de Apgar de 5 minuto | | | | | | | | | | | | | | | | | | | | | | | | |
| | Peso ao nascer em gramas | | | | | | | | | | | | | | | | | | | | | | | | |
| | Presença de malformação congênita | | | | | | | | | | | | | | | | | | | | | | | | |
| | Raça / cor da pele da mãe | | | | | | | | | | | | | | | | | | | | | | | | |
| | Número de gestações anteriores | | | | | | | | | | | | | | | | | | | | | | | | |
| | Número de partos normais (trabalhos de parto) | | | | | | | | | | | | | | | | | | | | | | | | |
| | Número de partos cesáreos (partos) | | | | | | | | | | | | | | | | | | | | | | | | |
| | Semana de gestação (por intervalos) | | | | | | | | | | | | | | | | | | | | | | | | |
| | Tipo de apresentação de recém-nascido | | | | | | | | | | | | | | | | | | | | | | | | |
| | Assistência ao parto | | | | | | | | | | | | | | | | | | | | | | | | |
| | Principal função do trabalhador | | | | | | | | | | | | | | | | | | | | | | | | |
| | Classificação do grupo Robson | | | | | | | | | | | | | | | | | | | | | | | | |
| | Morte antes de 28 dias (rótulo) | | | | | | | | | | | | | | | | | | | | | | | | |

Criação dos modelos:

- RFECV (Eliminação Recursiva de Feature com Validação Cruzada).
- Particionamento 90% e 10%.
- K-folds cross-validation.
- GridSearchCV.



Regressão Logística:
Utilizando as *features* selecionadas pelo RFECV

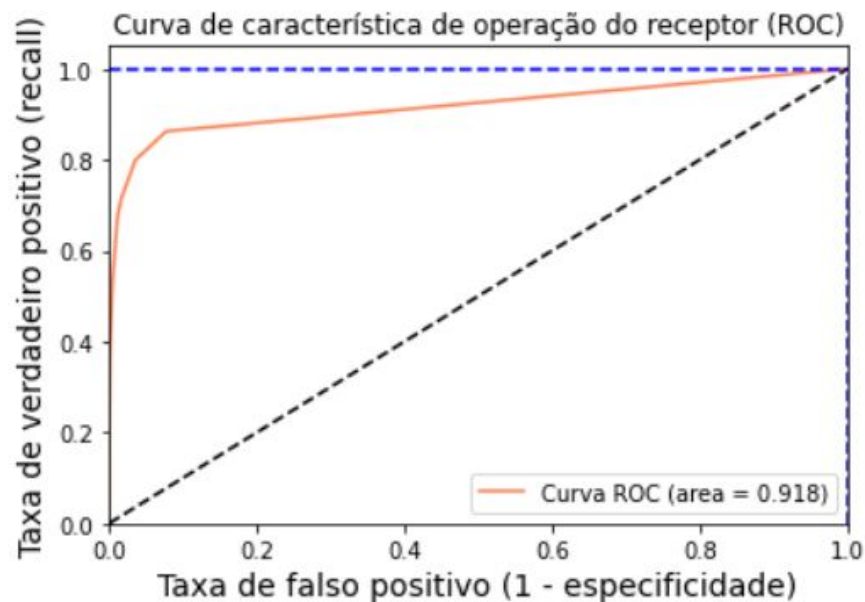


| | Resultados | | |
|----------|------------|--------------------------|--------------|
| | Padrão | K-Folds Cross Validation | GridSearchCV |
| Acurácia | 0.995 | 0.995 | 0.99 |
| AUC | 0.925 | 0.931 | 0.93 |
| Lod_Loss | 0.021 | 0.021 | 0.02 |

Regressão Logística:
Utilizando as *features* selecionadas pelo RFECV.

| | | Predito | |
|-------------|---------------------|-----------------------|------------------------|
| | | Nascidos vivos (0) | Nascidos mortos (1) |
| Classe Real | Nascidos vivos (0) | 671256 | 680 |
| | Nascidos mortos (1) | 2979 | 1108 |

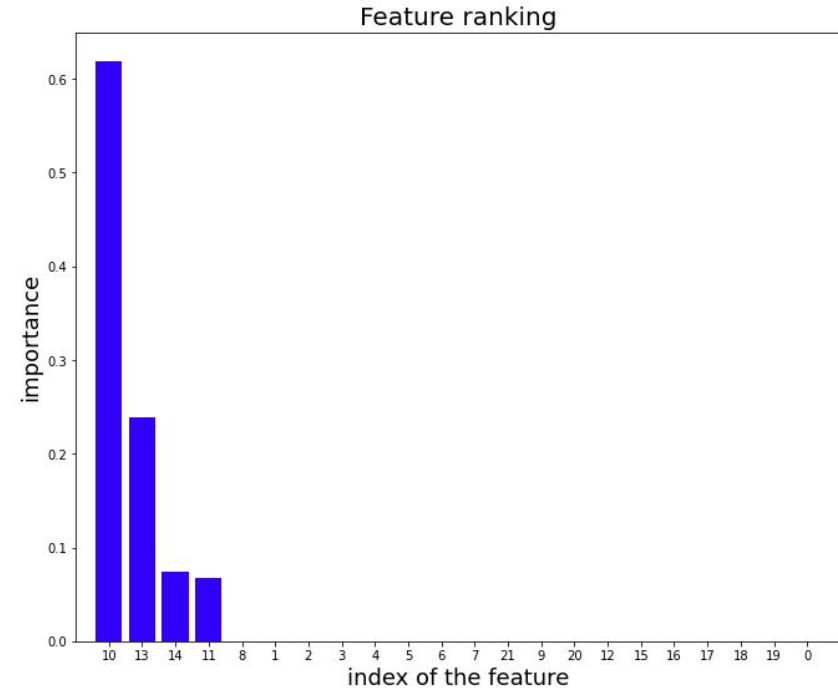
Árvore de Decisão :



| | Resultados | |
|----------|------------|--------------------------|
| | Padrão | K-Folds Cross Validation |
| Acurácia | 0.995 | 0.995 |
| AUC | 0.918 | 0.921 |
| Lod_Loss | 0.020 | 0.019 |

Árvore de Decisão :

| | | Predito | |
|-------------|---------------------|--------------------|---------------------|
| | | Nascidos vivos (0) | Nascidos mortos (1) |
| Classe Real | Nascidos vivos (0) | 670997 | 939 |
| | Nascidos mortos (1) | 2602 | 1485 |



- Avaliar o desempenho dos modelos pela acurácia não foi viável.
- A avaliação pela matriz de confusão foram melhores.
- Utilizar o GridSearchCV para melhorar as predições não foi funcional.
- Os mesmos algoritmos foram aplicados para a base de dados de São Paulo.
- O modelo de árvore de decisão se saiu melhor nas predições.

- CAMPOS, Raphael. **Árvores de Decisão**: Então diga-me, como construo uma?. [S. l.], 28 nov. 2017. Disponível em: <https://medium.com/machine-learning-beyond-deep-learning/%C3%A1rvores-de-decis%C3%A3o-3f52f6420b69>. Acesso em: 23 jan. 2021.
- BELUZO, Carlos Eduardo; SILVA, Everton; ALVES, Luciana Correia; BRESAN, Rodrigo Campos; ARRUDA, Natália Martins; SOVAT, Ricardo; CARVALHO, Tiago. **SPNeoDeath: A demographic and epidemiological dataset having infant, mother, prenatal care and childbirth data related to births and neonatal deaths in São Paulo city Brazil – 2012–2018**, [s. l.], 19 jul. 2020.
- LANSKY, S. et al. **Pesquisa Nascer no Brasil: perfil da mortalidade neonatal e avaliação da assistência à gestante** . Cadernos de Saúde Pública, Scielo, v. 30, p. S192 – S207, 00 2014.
- ILUMEO (Brasil). **Um tour pelos 10 principais algoritmos de Machine Learning**: Regressão Logística. [S. l.], 22 jun. 2020. Disponível em: <https://ilumeo.com.br/todos-posts/2020/06/22/um-tour-pelos-10-principais-algoritmos-de-machine-learning>. Acesso em: 23 jan. 2021.