

Portifólio Inteligência Artificial

A proposta deste trabalho visa proporcionar uma experiência prática na aplicação de técnicas de aprendizado de máquina, especificamente o uso de rede neurais, para a classificação de dados em um contexto específico, neste caso, a previsão da presença de diabetes.

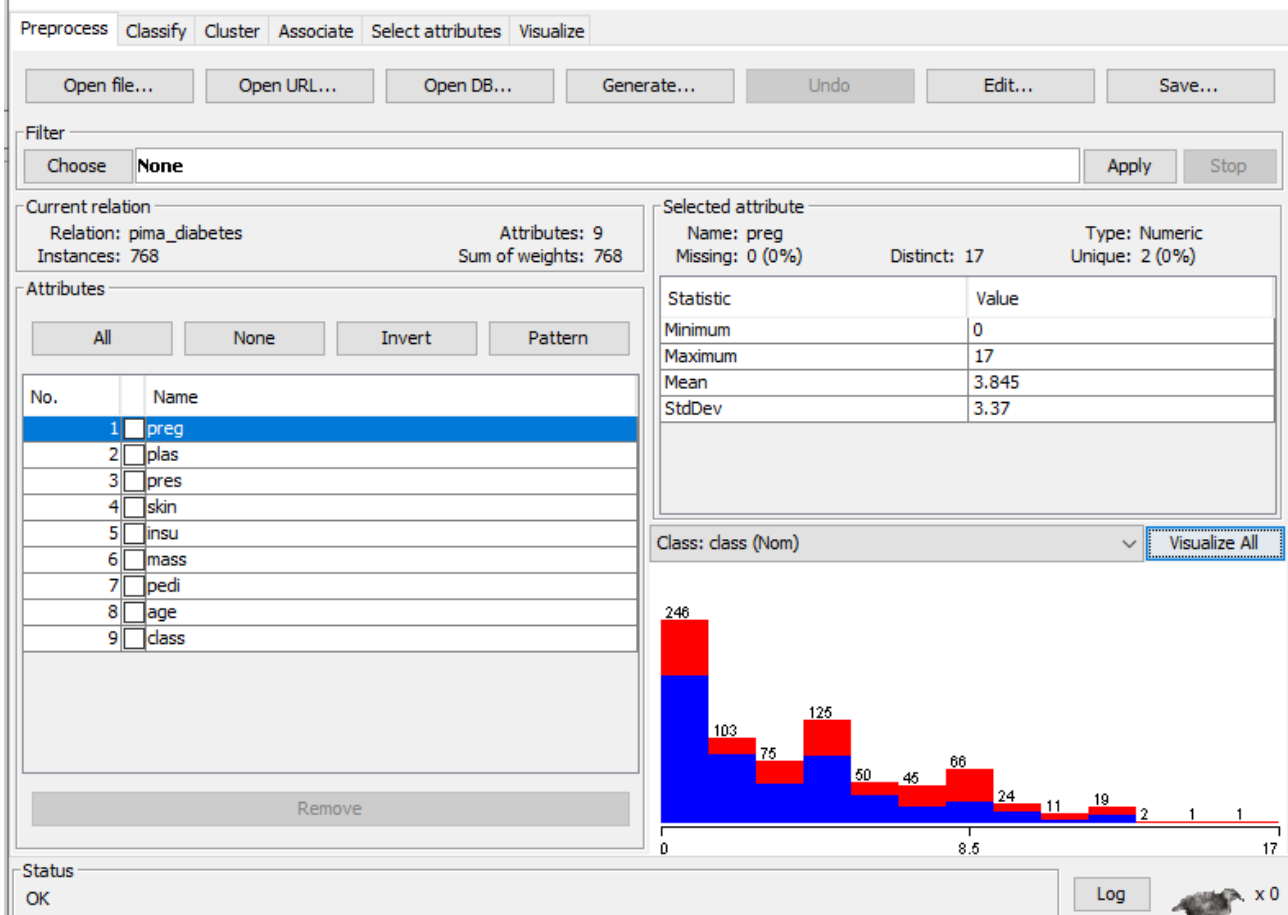
Aluno: Lucas de Carvalho Adam

Iniciando o Software

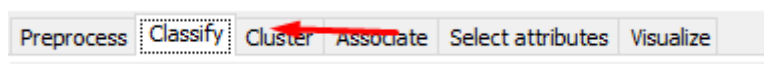
Primeiramente, busquei na internet tutoriais sobre como instalar o software Weka na minha máquina. Segui as instruções fornecidas para garantir uma instalação adequada.

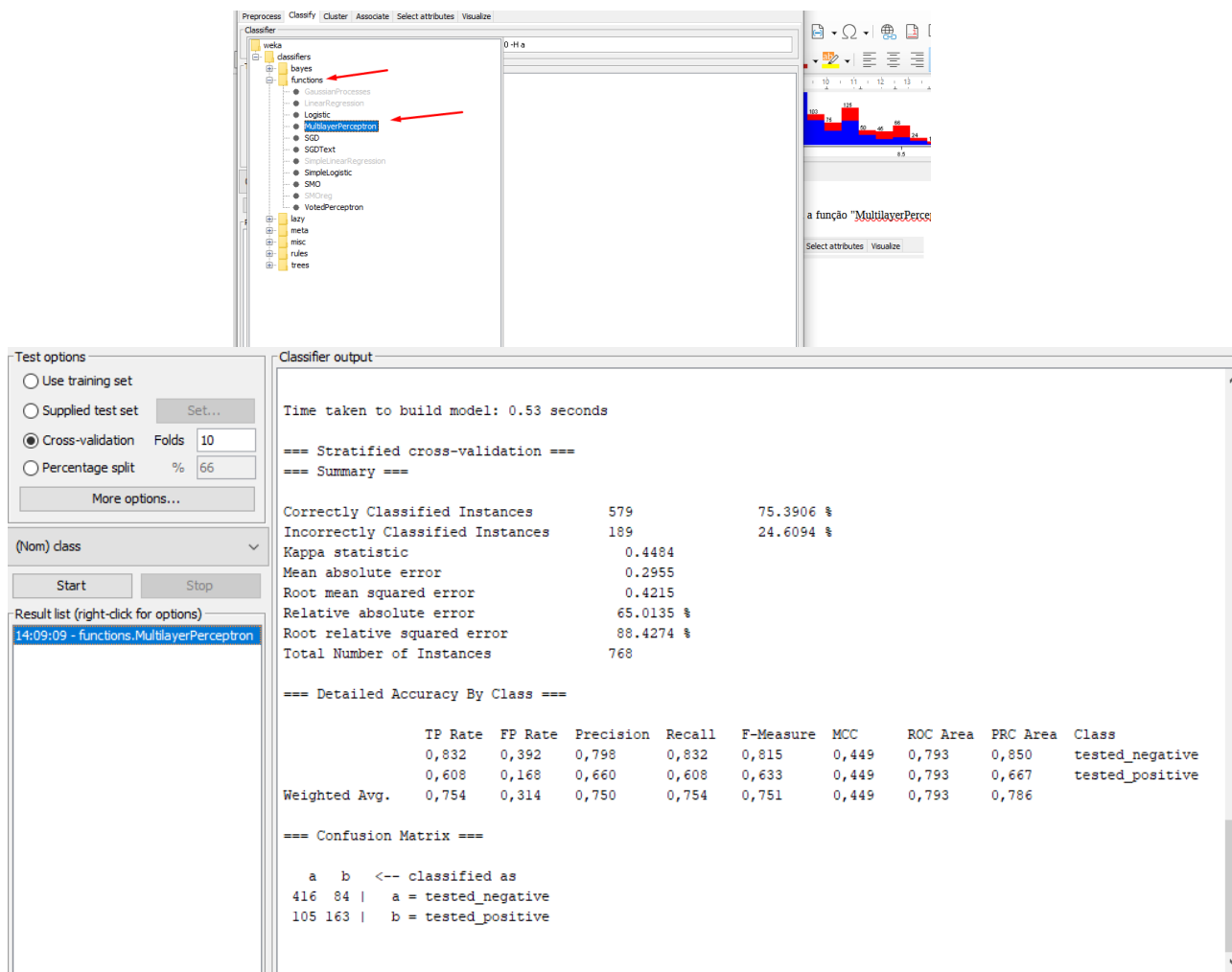
Implementação da Rede Neural para Previsão de Diabetes

Abri o Weka e carreguei a base de dados de diabetes na interface.



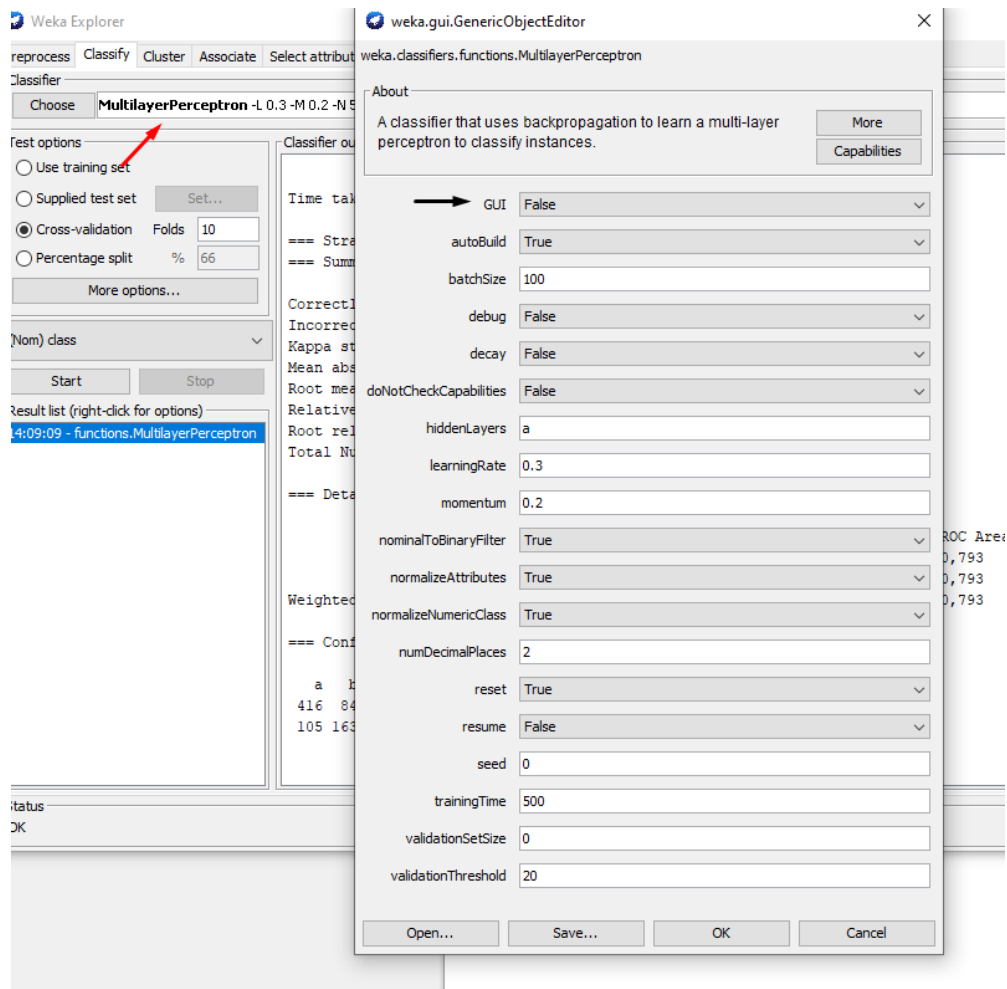
Na seção "Classify", cliquei em "Classifier" e escolhi a função "MultilayerPerceptron" como nosso classificador.



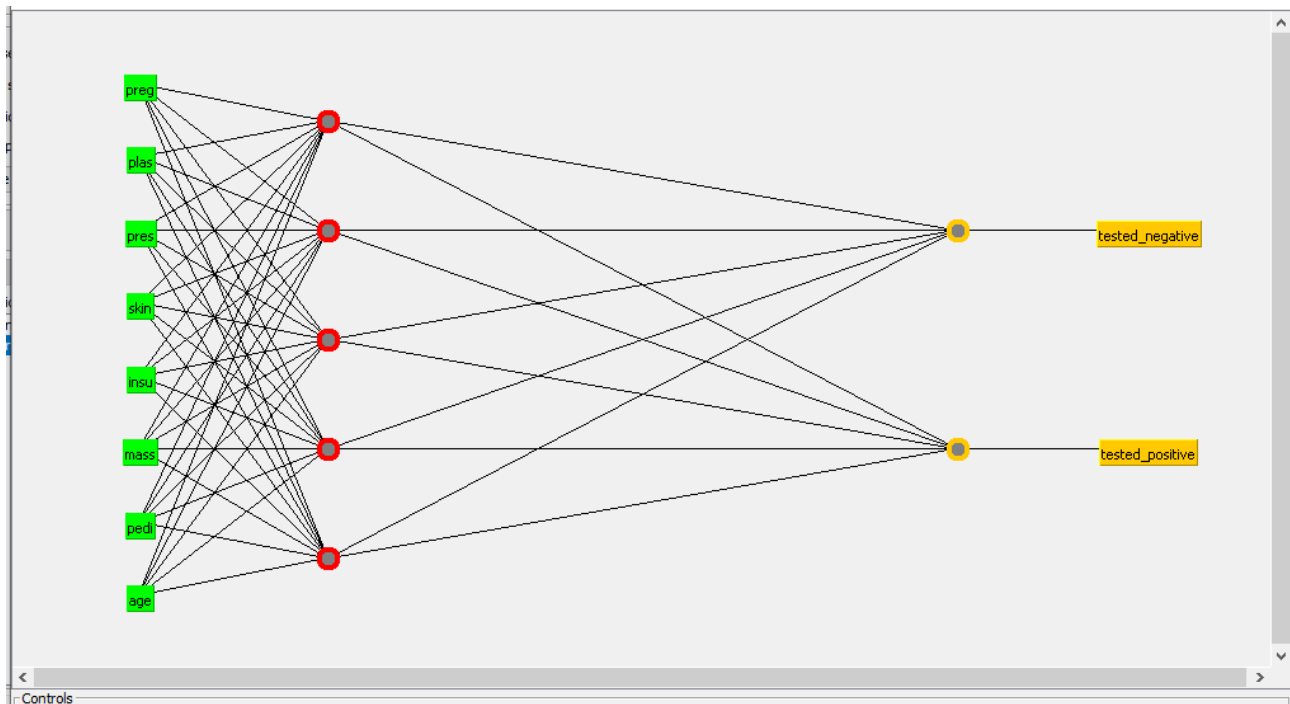


Configurei o modelo para usar o conjunto de dados de treinamento para prever a presença ou ausência de diabetes.

Habilitei a interface gráfica (GUI) no modelo "MultilayerPerceptron".

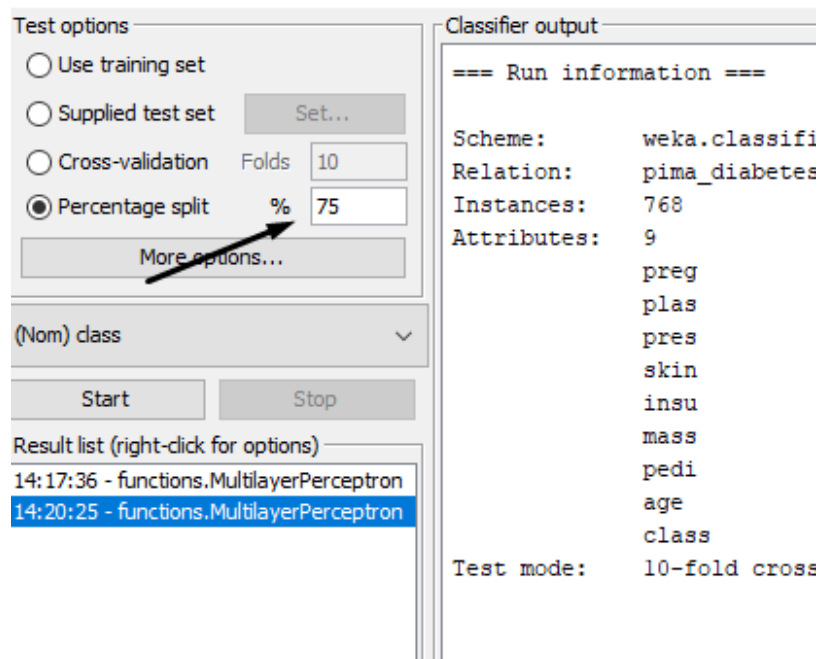


Executei o modelo e analisei os resultados, focando nas métricas de desempenho, incluindo o "Root mean squared error" e a "Confusion Matrix".



Divisão dos Dados

Para entender como a divisão dos dados afeta o desempenho do modelo, realizamos uma divisão, em que 75% dos dados foram designados para teste, enquanto o restante foi usado para treinamento.



Utilizamos a opção

"Percentage split" em "Test options" para realizar essa divisão.

Executamos o modelo novamente após a divisão dos dados.

```

=== Run information ===

Scheme:      weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R
Relation:    pima_diabetes
Instances:    768
Attributes:   9
              preg
              plas
              pres
              skin
              insu
              mass
              pedi
              age
              class
Test mode:    split 75.0% train, remainder test

=== Classifier model (full training set) ===

Sigmoid Node 0
  Inputs  Weights
  Threshold -2.7229322532147213
  Node 2    1.4723226659593016
  Node 3    7.820430499517925
  Node 4    2.4438687225161058
  Node 5    3.201530885671735
  Node 6    -3.028108946415126

Sigmoid Node 1
  Inputs  Weights
  Threshold 2.722932258743451
  Node 2    -1.4723226671089324
  Node 3    -7.820430653667111
  Node 4    -2.4438687292638233
  Node 5    -3.201530892432127
  Node 6    3.0281089587884145

Sigmoid Node 2

```

Compararamos detalhadamente os resultados da segunda execução com a primeira, levando em consideração as métricas de erro e a matriz de confusão.

A divisão para 75% durante o processo de treinamento do modelo no Weka, isso significa que 75% dos seus dados serão usados para treinamento e os restantes 25% serão usados para teste. A principal consequência dessa alteração é que haverá mais dados disponíveis para o treinamento do modelo, o que pode afetar a capacidade do modelo de aprender com mais precisão os padrões presentes nos dados.

Aqui estão algumas implicações ao aumentar a proporção de dados de treinamento (75%):

Maior Conjunto de Treinamento: Com 75% dos dados sendo usados para treinamento, o modelo terá um conjunto de treinamento maior. Isso pode permitir que o modelo aprenda com mais detalhes e nuances presentes nos dados, o que é geralmente benéfico.

Menos Dados de Teste: O conjunto de teste será menor, o que pode levar a uma avaliação menos precisa do desempenho do modelo. Isso ocorre porque o conjunto de teste é responsável por verificar quão bem o modelo generaliza para novos dados não vistos. Quanto menor o conjunto de teste, menos confiável é essa avaliação.

Possível Overfitting: Um conjunto de treinamento maior pode aumentar o risco de overfitting, que é quando o modelo se ajusta demais aos dados de treinamento e não generaliza bem para novos dados. Portanto, é importante monitorar o desempenho do modelo nos dados de teste para evitar overfitting.

Maior Tempo de Treinamento: Treinar um modelo com mais dados pode levar mais tempo computacional.

Esta atividade prática proporcionou uma compreensão valiosa da aplicação de redes neurais para prever diabetes usando o Weka. Aprendi a configurar o modelo, interpretar métricas de desempenho e percebi como a divisão dos dados pode impactar os resultados da previsão. Além disso, compreendi a importância de ajustar cuidadosamente as configurações do modelo e avaliar seu desempenho na previsão de uma condição crítica como o diabetes.

Lucas de Carvalho adam
Ponta Grossa- PR
2023