

Diagrama Geral

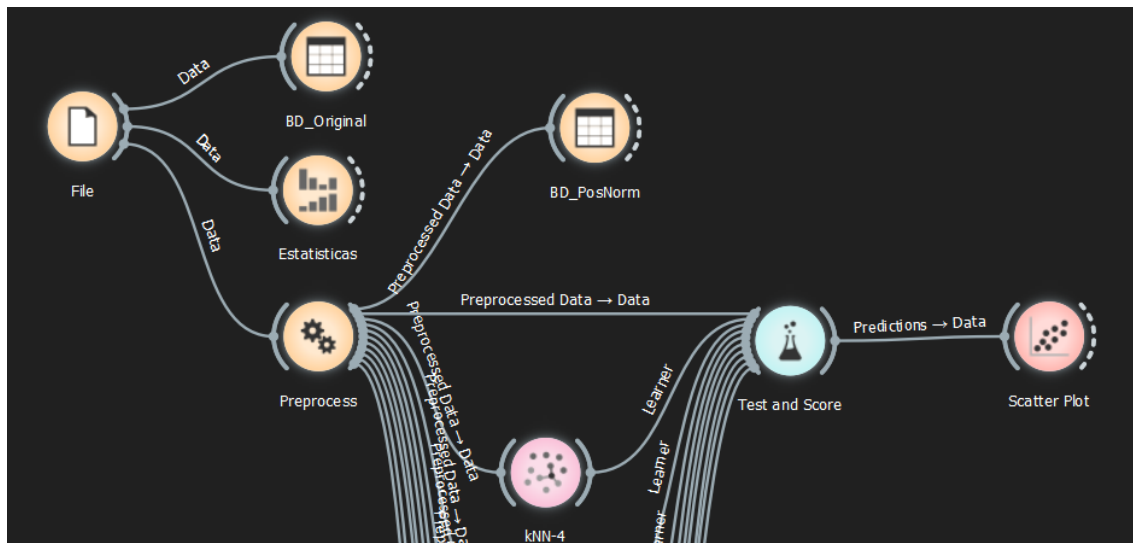
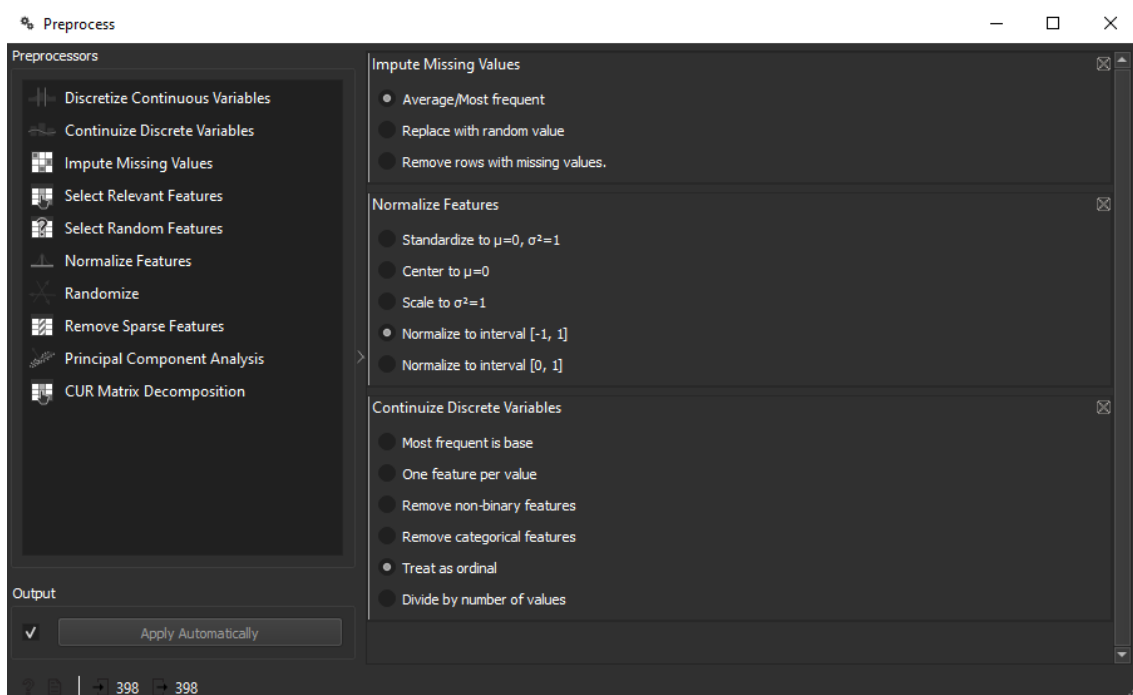


Diagrama do Processamento

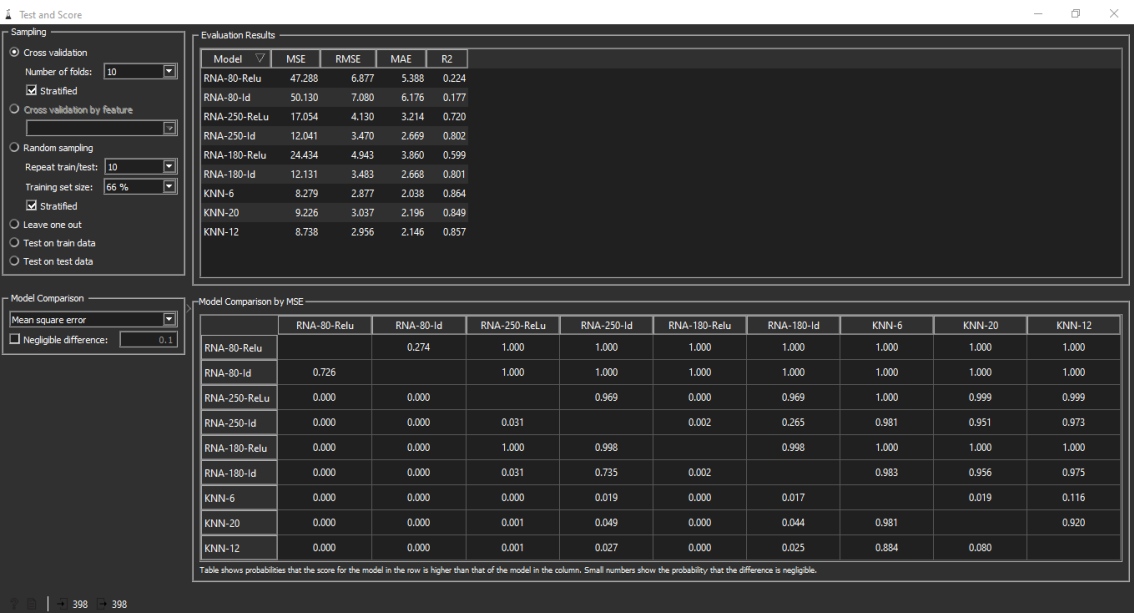
Foram analisadas as estatísticas da base para saber a necessidade de normalização. Utilizando o *Preprocess* os dados faltantes foram adicionados e a base foi normalizada. Os dados também foram “convertidos” para numéricos (*Treat as ordinal*).



Pré-processamento

Os testes com a base de dados foram feitos utilizando o kNN com 6, 12 e 20 vizinhos. As redes neurais testadas utilizaram as funções de ativação Identidade e Relu, com camada intermediária variando entre 80, 180 e 250.

Por ser uma base de dados Numérica, e não Categórica, não foi possível utilizar a matriz de confusão para obter uma avaliação mais direta dos erros.



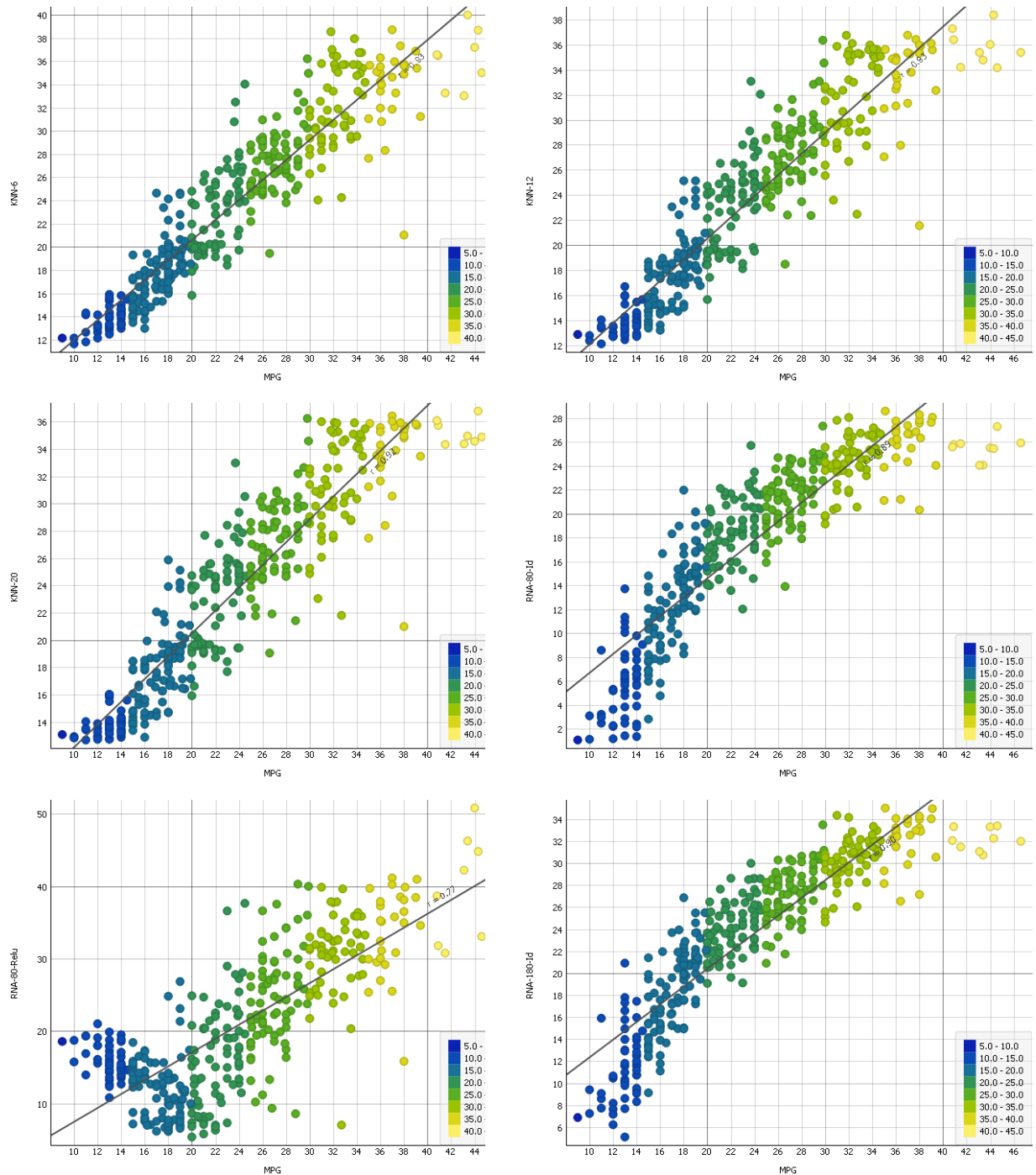
Test-Score obtido

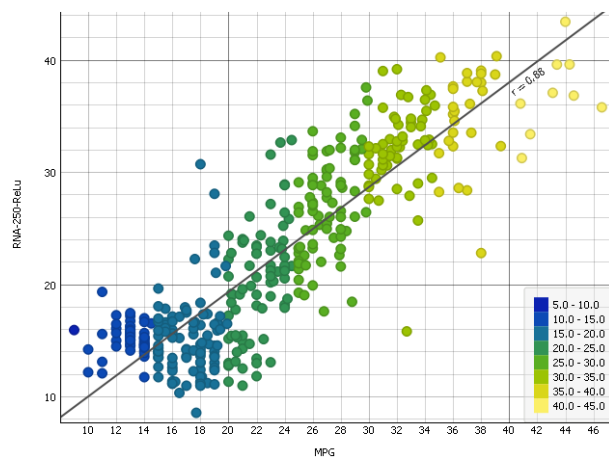
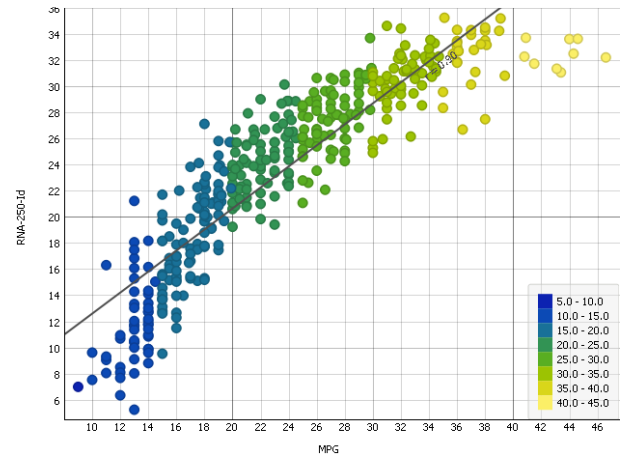
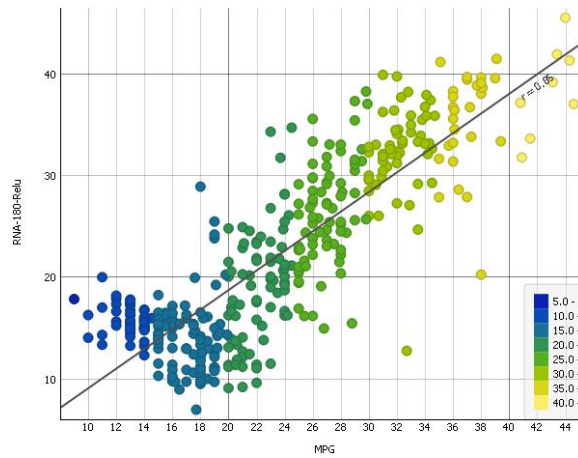
Model	MSE	RMSE	MAE	R2
RNA-80-Relu	47.288	6.877	5.388	0.224
RNA-80-Id	50.130	7.080	6.176	0.177
RNA-250-Relu	17.054	4.130	3.214	0.720
RNA-250-Id	12.041	3.470	2.669	0.802
RNA-180-Relu	24.434	4.943	3.860	0.599
RNA-180-Id	12.131	3.483	2.668	0.801
KNN-6	8.279	2.877	2.038	0.864
KNN-20	9.226	3.037	2.196	0.849
KNN-12	8.738	2.956	2.146	0.857

Média de erros dos testes

Analisando a média de erros absoluta percebe-se que o menor erro encontrado foi para o kNN com 6 vizinhos, e dentre as redes testadas, a Identidade com 180 camadas mostrou melhor resultado.

Analisando os gráficos de dispersão podemos ver visualmente qual dos testes teve mais pontos isolados (fora da linha geral dos valores de MPG). Estes pontos são os maiores erros das previsões.





Após análises foi possível perceber que os piores resultados estavam com kNN de 20 vizinhos, RNA Relu com 80 na camada intermediária, Relu com 180 e Relu com 50.

Em resumo, a RNA com função de ativação teve os piores resultados, e para a base de dados apresentada o melhor método é o kNN (com 6 vizinhos ou mais – sendo menos que 20).