

Relatório – Trabalho final usando Dataset para classificação de sementes de abobora

Disciplina: Data Science & IA

Especialização: Tech Lead - Cesar School

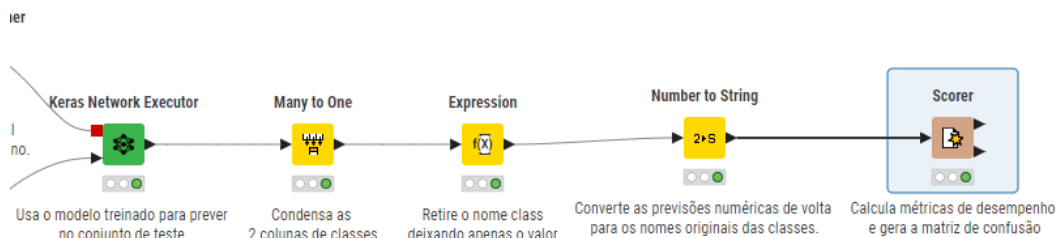
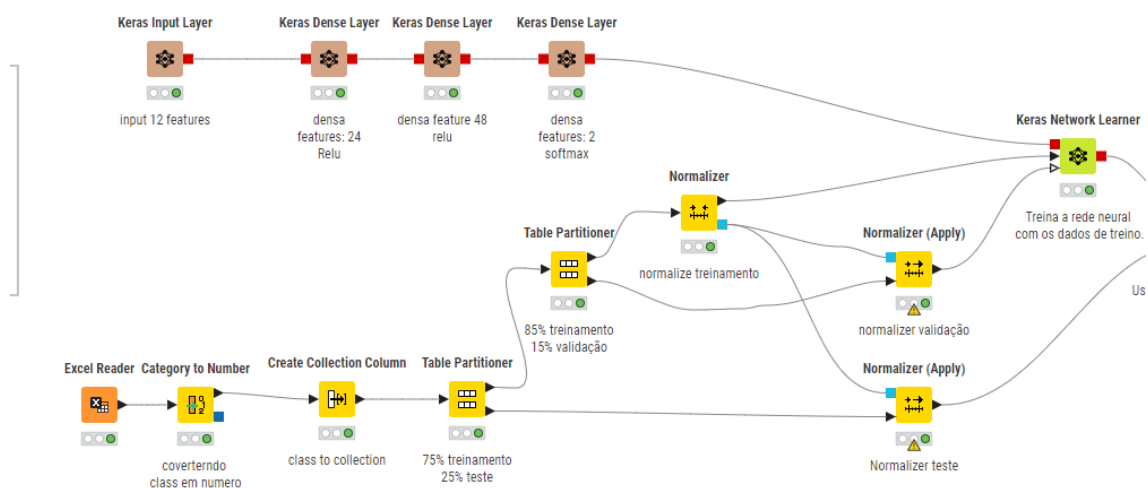
Aluno: Lucas Matos de Medeiros

email pessoal: lucas.matos.medeiros@gmail.com

email cesar: lmm5@cesar.school

Link do Dataset: <https://www.kaggle.com/datasets/muratkokludataset/pumpkin-seeds-dataset>

1. Imagem do modelo



2. Performance

2.1 matriz de confusão

► 1: Confusion matrix

► 2: Accuracy statistics

✕ Flow Variables

tows: 2

Columns: 2

Table

Statistics

<input type="checkbox"/>	#	RowID	0 Number (integer)	1 Number (integer)
<input type="checkbox"/>	1	0	300	25
<input type="checkbox"/>	2	1	51	249

Essa matriz de confusão mostra o desempenho do modelo na classificação das duas classes de sementes de abóbora:

Linha 0 (classe real 0)

300 amostras corretamente classificadas como 0 (*True Positives* para a classe 0).

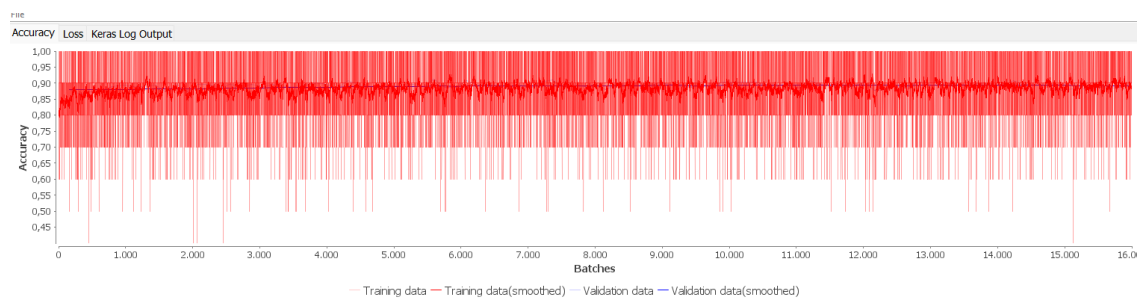
25 amostras incorretamente classificadas como 1 (*False Negatives* para a classe 0).

Linha 1 (classe real 1)

51 amostras incorretamente classificadas como 0 (*False Positives* para a classe 0).

249 amostras corretamente classificadas como 1 (*True Positives* para a classe 1).

2.2 Evolução do treinamento



- As curvas indicam **aprendizado consistente e convergência estável**.
- Não há sinais fortes de *underfitting* (modelo incapaz de aprender) nem *overfitting* severo.
- Pequena diferença entre treino e validação é normal e esperada, especialmente em problemas de classificação com dados reais.

3. Conclusão

A análise do modelo de classificação de sementes de abóbora demonstrou um desempenho satisfatório, com acurácia próxima de 88%. O modelo, baseado em rede neural com duas camadas densas e função de ativação ReLU, mostrou-se eficaz na diferenciação entre as variedades, utilizando principalmente atributos relacionados à forma e proporções da semente, como *Aspect Ratio*, excentricidade e compacidade. Como conclusão, o fluxo implementado no KNIME atende ao objetivo de classificar de forma automática e consistente, embora apresente margem para melhorias por meio de ajustes de hiperparâmetros, regularização e validação cruzada. Entre as principais dificuldades encontradas na construção do modelo no KNIME, destacam-se a correta configuração dos nós de pré-processamento para garantir normalização apenas nas variáveis preditoras, o mapeamento e tratamento adequado da coluna de classe para uso no Keras, a definição da arquitetura da rede neural dentro das limitações visuais da ferramenta, a integração dos nós de aprendizado e execução para manter compatibilidade de tipos e formatos de dados, e a interpretação dos resultados gerados pelo Scorer, especialmente quando se deseja extrair métricas específicas ou ajustar o fluxo para novas rodadas de teste.