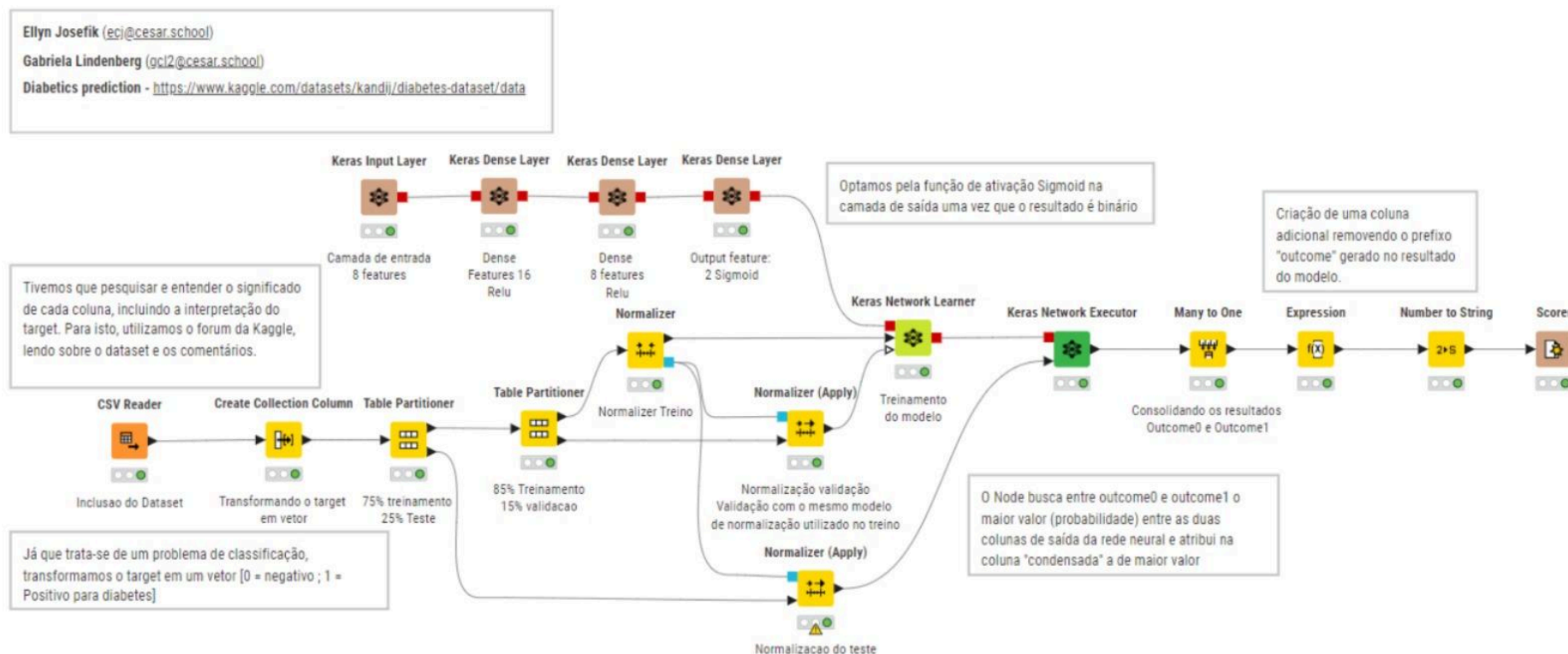


Ellyn Josefik - ecj@cesar.school

Gabriela Lindenberg - gcl2@cesar.school

Data set: <https://www.kaggle.com/datasets/kandij/diabetes-dataset/data>

KNIME Workflow



Análise de performance

Rows: 2 | Columns: 2

Table

Statistics

<input type="checkbox"/>	#	RowID	1 Number (Integer)	0 Number (Integer)
<input type="checkbox"/>	1	1	41	17
<input type="checkbox"/>	2	0	28	106

<input type="checkbox"/>	#	RowID	TruePositives Number (Integer)	FalsePositives Number (Integer)	TrueNegatives Number (Integer)	FalseNegativ... Number (Integer)	Recall Number (Float)	Precision Number (Float)	Sensitivity Number (Float)	Specificity Number (Float)	F-measure Number (Float)	Accuracy Number (Float)	Cohen's kappa Number (Float)
<input type="checkbox"/>	1	1	41	28	106	17	0.707	0.594	0.707	0.791	0.646	?	?
<input type="checkbox"/>	2	0	106	17	41	28	0.791	0.862	0.791	0.707	0.825	?	?
<input type="checkbox"/>	3	Overall	?	?	?	?	?	?	?	?	?	0.766	0.473

A rede neural treinada apresentou uma **acurácia de ~77%**, indicando que, de forma geral, o modelo acerta a classificação na maioria dos casos. Entretanto, ao observarmos a **matriz de confusão**, notamos que o modelo tem maior facilidade em identificar corretamente pacientes **sem diabetes** (classe 0), enquanto apresenta mais erros na detecção de pacientes **com diabetes** (classe 1). Essa tendência é confirmada pelas métricas:

- **Precisão (classe 1 – diabéticos):** o modelo não comete muitos falsos positivos, ou seja, quando classifica alguém como diabético, geralmente está correto.
- **Recall / Sensibilidade (classe 1 – diabéticos):** apresenta valores menores (~71%), o que significa que alguns pacientes diabéticos estão sendo classificados como não diabéticos (**falsos negativos**).
- **Especificidade (classe 0 – não diabéticos):** elevada, reforçando que o modelo tem bom desempenho em reconhecer indivíduos saudáveis.
- **F1-Score:** demonstra equilíbrio razoável entre precisão e recall, mas ainda indica espaço para melhorias na detecção da classe minoritária.

Esses resultados sugerem um **desbalanceamento de classes** no dataset, típico em problemas médicos, onde geralmente há mais exemplos de pacientes sem a condição do que com a condição. Isso pode levar o modelo a “preferir” acertar a classe majoritária.

No contexto de saúde, essa limitação é crítica:

- **Falsos negativos** (diabéticos classificados como não diabéticos) são mais graves, pois podem atrasar diagnósticos e comprometer o tratamento.
- **Falsos positivos** (não diabéticos classificados como diabéticos), embora indesejáveis, são menos prejudiciais, pois podem ser descartados em exames adicionais.

Portanto, ainda que a acurácia geral seja satisfatória, o modelo deve ser ajustado para **maximizar o recall/sensibilidade da classe positiva**, mesmo que isso implique uma leve queda na precisão ou na acurácia global.

Conclusão

Neste trabalho utilizamos o dataset público de predição de diabetes do Kaggle para construir um modelo de classificação binária no KNIME, explorando redes neurais com Keras. A arquitetura definida contou com 8 variáveis de entrada, duas camadas densas intermediárias com 16 e 8 neurônios ativados por ReLU, e uma camada de saída com função Sigmoid para classificar o paciente como positivo (1) ou negativo (0) para diabetes.

A acurácia obtida foi de aproximadamente **77%**, resultado satisfatório como ponto de partida. No entanto, ao analisarmos a matriz de confusão e métricas complementares, percebemos que o modelo tem melhor desempenho em identificar casos negativos (não diabéticos) do que positivos, o que pode ser explicado pelo possível **desbalanceamento de classes**.

Para problemas de saúde, métricas como **recall** e **sensibilidade** são críticas, pois falsos negativos (pacientes doentes classificados como saudáveis) podem gerar graves consequências. Nesse sentido, apesar do recall para casos positivos ter atingido ~71%, ainda existe espaço para melhorias, como:

- Ajustar hiperparâmetros da rede (número de neurônios, camadas, otimizadores e funções de ativação);
- Aplicar técnicas de balanceamento de classes (ex.: *SMOTE*, *oversampling* ou *class weights*);
- Explorar regularização (dropout, L2) para melhorar a generalização;
- Avaliar outras métricas de forma prioritária, como recall e f-measure, além da acurácia.

Concluimos que o modelo implementado já fornece resultados consistentes, mas que sua aplicação prática em saúde exigiria ajustes para privilegiar a detecção correta de casos positivos, reduzindo os riscos associados a falsos negativos.