

Disciplina de Inteligência Artificial

Relatório de análise do modelo

Alunos: Igor Schokal Schons

Professor: Diego Pinheiro

Novo Hamburgo, 26/06/2025

Objetivo: Prever se um tumor de mama é benigno ou maligno. Com base em aspectos e características de cada tipo de tumor, como raio, textura, perímetro, área, suavidade, entre outros.

Descrição e justificativa: Escolhi esse assunto por conta da sua importância no meio da saúde, esse projeto futuramente pode ser de grande valia para estudos clínicos avançados. O dataset é tem 569 registros e 31 colunas, as colunas são compostas por: diagnosis, radius_mean, texture_mean, perimeter_mean, area_mean, smoothness_mean, compactness_mean, concavity_mean, concave_points_mean, symmetry_mean, fractal_dimension_mean, radius_se, texture_se, area_se, smoothness_se, compactness_se, concavity_se, concave_points_se, symmetry_se, fractal_dimension_se, radius_worst, texture_worst, perimeter_worst, area_worst, smoothness_worst, compactness_worst, concavity_worst, concave_points_worst, symmetry_worst, fractal_dimension_worst. Devemos nos atentar que o modelo foi treinado apenas com os valores médios, que são representados pelas colunas que contém o “_mean”.

Análise: Há uma grande variedade de atributos sobre os tipos de tumores no dataset, por isso a decisão de usar apenas os valores médios no treinamento do modelo. A variável “diagnostics” foi alterada para conter um valor binário, onde:

- True representa tumor maligno, “M”.
- False representa tumor benigno, “B”.

Pré-processamento realizado:

- Seleção das colunas que contém os valores médios.
- Criação da variável alvo binária.
- Divisão do dataset em treino (70%) e teste (30%).

Descrição do Modelo Escolhido e Processo de Treinamento:

Para prevermos se um tumor é benigno ou maligno utilizamos um modelo de rede neural artificial do tipo sequencial, que é composto por camadas densamente conectadas. O modelo possui uma camada de entrada com 64 neurônios e utiliza a função de ativação Relu, na sequência temos um camada de dropout de 50%, essa camada de dropout se faz necessária para evitar o overfitting no nosso modelo. Em seguida, foi adicionada outra camada com 32 neurônios e uma nova camada de dropout com 30%. A camada final possui apenas um neurônio com a função de ativação Sigmoid, que é necessária para as tarefas de classificação binária. O modelo foi compilado com o otimizador Adam, ele utiliza a função de perda binary_crossentropy. A métrica de avaliação do nosso modelo é a acurácia. O

processo de treinamento foi feito por até 400 épocas. Utilizamos o EarlyStopping para interromper o treinamento do modelo caso começasse a ocorrer uma perda de qualidade após 10 épocas. O conjunto de testes foi carregado com 30% dos dados, enquanto o de treinamento foi carregado com 70%.

Resultados e análise da performance do modelo:

Os resultados mostraram que o modelo tem uma performance bem satisfatória. A acurácia atingiu 98%, o que indica uma boa capacidade de prever corretamente o diagnóstico dos tumores. Além disso, a curva de perda de validação apresentou uma queda rápida nas primeiras épocas, o que deixa claro que o modelo conseguiu aprender de forma eficiente sem sofrer com o overfitting. Porém o uso de técnicas de regularização como Dropout e EarlyStopping contribuíram significativamente para esse desempenho do modelo, garantindo uma boa generalização dos dados. Com base nesses fatos podemos concluir que o modelo se mostra confiável.