

[< Return to "Machine Learning Engineer Nanodegree" in the classroom](#)

Machine Learning Capstone Project

REVISÃO

REVISÃO DE CÓDIGO

HISTORY

Meets Specifications

Parabéns pelo trabalho! O relatório ficou muito bom e acredito que servirá de base para futuros trabalhos como engenheiro de machine learning. 🎉

Espero ter contribuído com seu aprendizado. Desejo a você uma carreira de sucesso! 😊

Definição do problema

O relatório contém uma explicação em alto nível do projeto em termos leigos, com informações sobre o domínio do problema, sua origem e o conjunto de dados utilizado.

A introdução descreve de maneira concisa o problema enfrentado.

O problema abordado é definido com clareza. Uma estratégia para resolvê-lo e a solução esperada são discutidos.

O problema foi bem definido junto com a estratégia para resolução.

As métricas utilizadas para medir o desempenho do modelo ou seus resultados são claramente definidas. A escolha de métricas é justificada com base na caracterização do projeto.

Excelente a utilização do parâmetro beta para ajustar o peso da precisão no cálculo da métrica F1! 🎉

Análise do problema

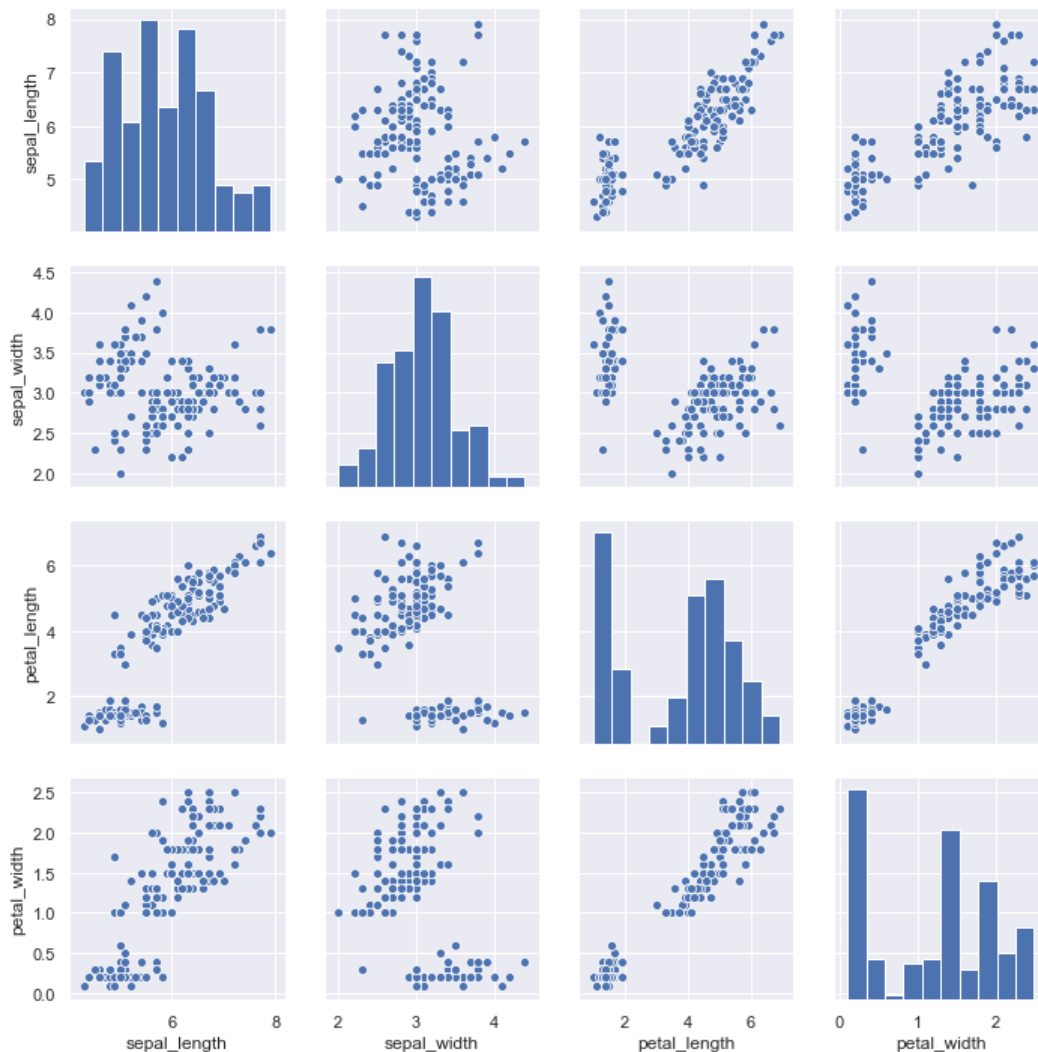
Se um conjunto de dados está disponível, atributos e estatísticas relevantes para o problema são calculados, relatados e discutidos, acompanhando uma amostra dos dados. Na ausência de um conjunto de dados, uma descrição detalhada do domínio matemático das entradas é feita. Anormalidades ou características dos dados que precisam ser tratadas são identificadas.

Foi utilizado um conjunto de dados público, exibidas algumas entradas e as features foram descritas com um dicionário de dados.

Uma visualização que sumariza ou extrai uma característica ou atributo relevante do conjunto de dados é criada e discutida. Indicadores visuais desse atributo são definidos com clareza.

Foram incluídos histogramas e gráfico que barras demonstrando a distribuição da variável alvo.

Poderiam também ser incluídos histogramas e gráficos de distribuição das outras variáveis. O [pairplot](#) pode ser utilizado para avaliar a distribuição das variáveis:



Os algoritmos e técnicas utilizados no projeto são discutidos em detalhes e devidamente justificados com base nas características do problema.

É utilizado um modelo de referência claro com os quais se comparam os resultados e o desempenho das soluções desenvolvidas.

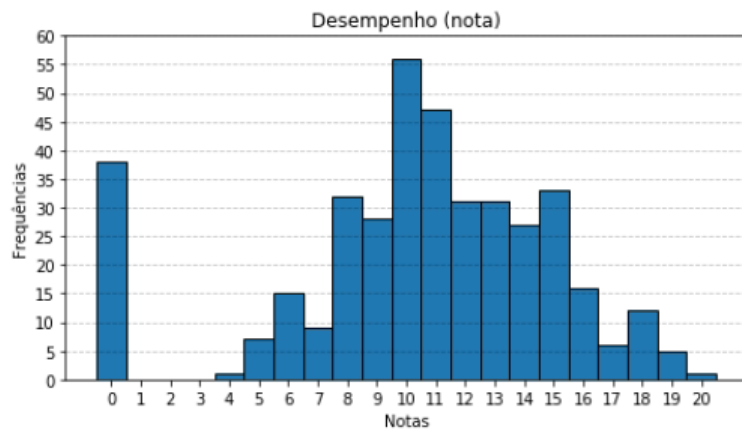
Muito bom! Foi utilizado um trabalho acadêmico como referência.

Metodologia

Todos os passos de pré-processamento de dados são documentados com clareza. Anormalidades ou características dos dados são ajustadas conforme necessário. Se não for necessário nenhum pré-processamento, isso é justificado detalhadamente.

Os outliers podem ser um problema para o aprendizado do modelo.

Os alunos com nota zero podem ser considerados outliers nesse caso pois podem ser bons alunos que por um motivo externo não fizeram a avaliação. Perceba que temos uma grande frequência de alunos com nota zero:



Nesse caso, sugiro que experimente utilizar alguma técnica para o tratamento de outliers.

Veja esse [post](#) sobre as técnicas para tratar dados ausentes, discrepantes e derivados.

O processo de implementação das métricas, algoritmos e técnicas utilizados com o conjunto de dados é detalhadamente documentado. Dificuldades ocorridas durante a implementação são documentadas e discutidas.

O processo de aprimoramento dos algoritmos e técnicas utilizados é claramente documentado. Os resultados iniciais e finais são relatados, além de passos intermediários quando cabível.

A otimização dos parâmetros do modelo é uma etapa que pode melhorar significativamente os resultados.

A maioria dos modelos são parametrizáveis e a escolha dos melhores parâmetros é importante para ajustar o modelo ao problema.

Por exemplo, no caso do SVM foi utilizado apenas o kernel linear mas só em problemas bem simples conseguimos resolver com um modelo linear. Existem outros tipos de kernel que costumam gerar melhores resultados. Sugiro fortemente que otimize os parâmetros dos modelos para encontrar melhores resultados.

Resultados

O modelo final, incluído seus parâmetros, é avaliado em detalhes. É feita uma análise para validar a robustez da solução encontrada.

Seria importante reportar também o desvio padrão dos resultados. Isso é essencial para a avaliar a variação dos resultados encontrados.

Os resultados finais são comparados com o modelo de referência escolhido através de algum método estatístico. A conclusão de que o modelo desenvolvido é suficiente para resolver adequadamente o problema é justificada.

O modelo superou o benchmark selecionado.

Ainda assim, podemos perceber que apenas 31% dos alunos que precisam de reforço foram identificados. Como foi dado um peso maior para a precisão, a revocação acabou sacrificada.

Conclusão

Foi criada uma visualização enfatizando uma qualidade importante do projeto, acompanhada de discussão detalhada. Indicativos visuais são claramente definidos.

A solução do problema de ponta a ponta é resumidamente discutida, destacando aspectos particularmente interessantes ou desafiadores do projeto.

O relatório discute como pelo menos um aspecto da implementação que possa ser aperfeiçoado. O resultado potencial desse aperfeiçoamento é comparado ou contrastado com a solução atual.

Diversos pontos de melhorias foram indicados.

Qualidade

O relatório do projeto segue uma estrutura bem organizada e é compreensível para sua audiência. Cada seção é escrita de maneira clara, concisa e específica. O texto não contém erros gramaticais ou ortográficos em excesso. Todas as referências utilizadas na confecção do projeto são citadas.

O código é bem formatado e contém comentários para explicar implementações completas. O output do código reflete os resultados discutidos no relatório.

 [BAIXAR PROJETO](#)

[RETORNAR](#)