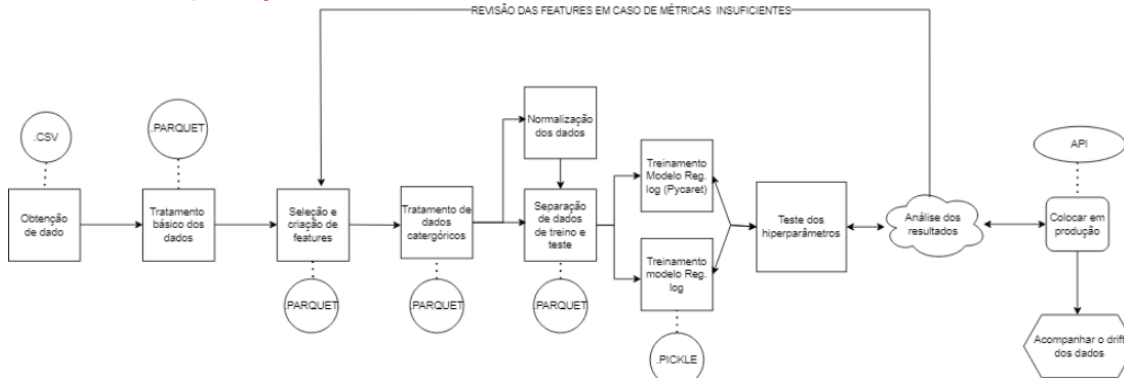


NOME: GABRIEL FONSECA RODRIGUES

2. Para começar o desenvolvimento, desenhe um diagrama que demonstra todas as etapas necessárias em um projeto de inteligência artificial desde a aquisição de dados, passando pela criação dos modelos, indo até a operação do modelo.



3. Descreva a importância de implementar pipelines de desenvolvimento e produção numa solução de aprendizado de máquinas.

Os pipelines são importantes no desenvolvimento de modelos de machine learning, pois dividem um processo complexo em várias tarefas mais simples e individualizadas. Dessa forma, o código se torna modular, ou seja, caso seja necessário alterar algo em alguma das etapas do processo a modificação será somente naquela etapa (Nó) específico, não afetando o restante do código. A principal vantagem da criação do pipeline é permitir testes acrescentando ou retirando etapas com poucas alterações no código.

Além disso, mantém o código mais organizado, portanto, caso outro desenvolvedor precise alterar o código não terá grande dificuldade para entender o processo.

4. Como as ferramentas Streamlit, MLFlow, PyCaret e Scikit-Learn auxiliam na construção dos pipelines descritos anteriormente? A resposta deve abranger os seguintes aspectos:

Rastreamento dos experimentos: O MLFlow facilita muito o processo de rastreamento dos modelos, pois permite o acompanhamento das métricas de diferentes modelos, salvando os hiperparâmetros, tempo de execução, função utilizada, entre outras variáveis de cada vez que o pipeline é executado. Dessa forma fica fácil rastrear qual a combinação de todas essas variáveis resultou no melhor modelo

Funções de Treinamento: Tanto o Scikit-Learn, como o PyCaret são de grande valia durante o processo de análise das funções de treinamento. O Scikit tem diversas funções prontas que permite ao desenvolvedor testar e analisar os resultados de acordo com o seu conjunto de dados. Já o PyCaret é uma ferramenta que testa diversos modelos (e diversos conjuntos de hiperparâmetros) para um mesmo conjunto de dados e retorna o que apresentar a melhor métrica. De forma geral o PyCaret é uma ótima forma de fazer uma análise inicial, onde o estudo do modelo retornado poderá ser aprofundado utilizando as funções do Scikit-Learn.

Monitoramento da saúde do modelo: O monitoramento do modelo pode ser realizado tanto com o rastreamento das métricas do MLFlow, analisando se as distribuições dos dados atuais apresentam resultados dentro do intervalo encontrado durante o processo de treino

Atualização do modelo: A atualização do modelo poderá ser realizada pelo Pycaret ou pelo Scikit Learn, pois caso os dados atuais tenham um drift muito significativo em relação aos dados que foram utilizados para treinar o modelo, será necessária uma nova investigação das features e um retreinamento do modelo

Provisionamento: A ferramenta Stremlit auxilia no provisionamento do modelo, pois permite a criação e o compartilhamento de modelos em API sem a necessidade de conhecimento prévio em programação front-end