Aplicação dos conceitos de Engenharia de Aprendizado de Máquina em Produção

RESUMO EXPANDIDO - Disciplina de TCC29009

André Luiz Faraco Mazucheli
Estudante do Curso de Engenharia de Telecomunicações

Mário de Noronha Neto, Dr.
Professor orientador

Semestre 2022.2

Resumo- *Com o avanço da tecnologia, cada vez mais se torna evidente o seu impacto em nosso cotidiano. Uma que se destaca cada vez mais é o Machine Learning (ML). Em seu montante de aplicações, existem muitos desafios a serem enfrentados quando se trata da engenharia dos dados que são utilizados em cada sistema, e também muitos obstáculos que surgem quando ele é colocado em produção. Estima-se que finalizar a modelagem e definição dos dados, pode representar apenas metade do que seria a primeira versão estável, confiável e eficiente do sistema. Este trabalho visa aplicar conceitos de Machine Learning Operations (MLOPs), em cenário de produção, explorando cada etapa do ciclo de vida de um sistema de ML, com o intuito de definir boas práticas e ferramentas para auxiliar no processo.*

Palavras-chave: Machine Learning. MLOPS. Engenharia de dados.

1 Introdução

Sistemas de Aprendizado de máquina, do inglês *Machine Learning*(ML), possuem um ciclo de vida, que tende a aprimorar cada vez mais sua eficiência. O ciclo pode ser dividido em 4 estágios básicos, sendo eles: **Escopo**, **Dados**, **Modelagem** e **Implantação**, podendo ser os 3 últimos, iterativos entre si. Cada etapa, possui uma série de processos e análises, que conforme são aperfeiçoados, melhoram a precisão e o desempenho do sistema, que são suscetível a mudanças com a variação do tempo em ambiente de produção.

Dentro deste contexto, podemos trabalhar com o conceito de Engenharia de Aprendizado de Máquina, que também pode ser chamado de **MLOps** (*Machine Learning Operations*). Ele agrega um conjunto de ferramentas e princípios para apoiar o progresso ao longo do ciclo de vida de um projeto de ***ML***, principalmente referente às etapas de Dados,

Modelagem e Implantação. A ideia central em MLOps é encontrar maneiras de pensar sobre o escopo de modelagem e implantação de dados e também ferramentas de software para sustentar as melhores práticas (ASHMORE; CALINESCU; PATERSON, 2021).

A importância deste conceito, está no fato de que quando um algoritmo de predição é utilizado em um sistema de ***ML***, ele ainda precisa passar por um processo de amadurecimento em produção, sendo necessário através de um processo iterativo do algoritmo, ser retreinado e ajustado para eventuais alteração em parâmetros não considerados previamente. Porém, este processo por sua vez, não é gerido e nem implementado muitas vezes de forma coerente.

Outro ponto de muito impacto no desempenho de um sistema ***ML***, é a questão do processo de implantação. Alguns pontos são extremamente importantes na hora de modelar sua estrutura. Definir bem se seu funcionamento ocorrerá em tempo real ou não, ou se ele será executado em nuvem, ou em *edge*, com o hardware e software em algum servidor local, e neste ponto é relevante considerar questões como latência de comunicação, que de fato pode influenciar muito. Questões de segurança de dados, também devem ser consideradas relevantes, já que o sistema por si só, trabalha com a análise de dados(NG, 2022).

1.1 Objetivo Geral

Baseado neste contexto, este trabalho possui o objetivo geral de aplicar conceitos de Engenharia de Aprendizado de Máquina em produção, utilizando um modelo de detecção de anomalias, definindo boas práticas para gerenciar e realizar ajustes em toda a infraestrutura definida.

1.2 Objetivo específicos

Visando atingir o objetivo geral, os seguintes objetivos específicos deverão se aplicados:

1. Pesquisar e entender os conceitos de MLOps;
2. Descrever o ciclo de vida de um sistema de ML em produção;
3. Criar a infraestrutura para implantação do sistema;
4. Definir aplicação e base de dados utilizada;
5. Gerar o modelo de detecção de anomalias;
6. Colocar o modelo em um cenário de produção;

2 Metodologia

Os tópicos abordados abaixo possuem a finalidade de fundamentar e explorar os conceitos e tecnologias utilizadas neste trabalho, para contextualizar a ideia e consolidar as motivações presentes nesta aplicação.

2.1 Componentes de um sistema de Aprendizado de Máquina

Primeiramente, se faz necessária toda a contextualização de como é composta a infraestrutura de um sistema de aprendizado de máquina. O objetivo é promover a consolidação da ideia de que no desenvolvimento de um sistema de *Machine Learning*, o algoritmo utilizado apesar de fundamental, não contempla toda o embasamento e suporte que o sistema necessita para ser otimizado, e se perpetuar em produção.

A definição de fato é que, assim que o sistema é desenvolvido e inicialmente parametrizado, após ser implantado em produção, se inicia o ciclo de Engenharia de dados através da aplicação de seus conceitos e boas práticas.

2.2 Ciclo de vida de um sistema de Machine Learning em produção

O ciclo de vida de um sistema de ML, é um processo iterativo e complexo que começa com a coleta dos dados usados para treinar um algoritmo de ML para um sistema e termina com a implantação desse componente dentro do sistema em produção (ASHMORE; CALINESCU; PATERSON, 2021). A contextualização detalhada de todo o ciclo é fundamental para consolidação dos conceitos aplicados neste trabalho.

Ele é iniciado pelo **Escopo do projeto**, esta etapa tem por objetivo a especificação e o planejamento do projeto. A maior importância nela é realizar questionamentos fundamentais para o seu desenvolvimento.

Questões como **"Quais são os recursos em termos de (dados, tempo, pessoas) necessários para o desenvolvimento do projeto?**, ou **"Quais são as métricas de sucesso do projeto?"**. Desta forma, o processo de **Escopo do projeto**, esta diretamente vinculado a procurar um problema de negócio para resolver com Inteligência Artificial (IA), sempre visando analisar a viabilidade técnica e valor.

Na sequência, o ciclo segue através da etapa de **Dados**, que engloba deste a forma como deve ser feita a definição dos dados que deseja capturar do cenário de aplicação, até a parte de tratativa destes dados. Estas parametrizações variam de acordo com cada caso, e é importante ressaltar que quanto mais organizados e rotulados os dados estiverem, maior será a otimização do sistema.

A etapa de **Modelagem** é responsável por treinar o algoritmo, ou concluir que é necessário a inserção de mais dados no sistema, indicando que retorne para a etapa anterior. Este processo iterativo de entre os dados e a modelagem é extremamente influente no resultado final do sistema a ser desenvolvido.

Por fim, a etapa de **Implantação**, que é de fato quando o sistema é implantado em produção, e é possível analisar da forma mais empírica se existe a necessidade de mais ajustes nas etapas anteriores do ciclo de vida.

Um dos problemas mais impactantes nesta etapa é a parte de Problemas estatísticos, que estão vinculados ao fato dos dados de entrada do sistema que treinaram o algoritmo, sofrerem variações com o tempo, e estas variações podem aparecer em dois cenários distintos, sendo eles uma mudança gradual, que seria uma mudança suave ao longo de um período de tempo, ou uma mudança repentina, que ocorre de forma brusca e pode acarretar em algum problema. A solução para este caso é alimentar o sistema com novos dados

e retreinar o algoritmo (NG, 2022).

2.3 Ferramentas para implantação

Serão utilizados neste trabalho, ferramentas de apoio para facilitar e robustecer a implementação do sistema. Para a infraestrutura, será utilizado o ***Docker***, com o intuito de facilitar a implantação nos mais diversos cenários de produção, e também tornar o sistema escalável.

Para implantar o pipeline de ML de produção, será utilizado o *TensorFlowX* (TFX), que é uma plataforma em escala de produção do Google utilizada por diversas empresas contempladas no mercado, como *Spotify*, *Airbus* e *Gmail*. Seu uso será importante na criação da estrutura e no gerenciamento das bibliotecas compartilhadas para integrar os componentes necessários para definir, iniciar e monitorar o sistema de aprendizado de máquina desenvolvido.

Para realizar interações com o pipeline do sistema será utilizado o *framework* Fast API. A ferramenta é focado no desenvolvimento de API's de alto desempenho, altamente utilizada em cenários que necessitam de desempenho e robustez.

3 Considerações Parciais

Com os pontos abordados neste trabalho, utilizando um modelo de detecção de anomalias, espera-se ilustrar que desenvolver o algoritmo de ***ML*** em um projeto, é apenas uma etapa, sendo estes processo envolvido por uma série de questões mais complexas que circundam o trabalho que serão abordadas. Após definir o treinamento do ***ML***, para obter o máximo de valor agregado no modelo de treinamento, sempre deve-se pensar no cenário em produção, ter noção de todo o ciclo de aprendizado de máquina do modelo de treinamento para realizar ajustes e definir critérios de parametrização do sistema, para gerenciar de forma eficiente todo o projeto.

Referências

ASHMORE, R.; CALINESCU, R.; PATERSON, C. Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 54, n. 5, may 2021. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/3453444>.

NG, A. *Machine Learning Engineering for Production (MLOps)*. 2022. Urlhttps://www.coursera.org/specializations/machine-learning-engineering-for-production-mlops?utm$_campaign = video-youtube-mlops-video-seriesutm_medium = institutionsutm_source = deeplearning - ai.$