



Aprendizado de Máquina – Grupo: 1

Semestre: 2024.1

Nome: Fause Carlos

Matrícula: 211031691

Lorenzo de Lima

190032821

Pedro Cabeceira

211043727

Rafael Kenji

190044128

Artur Jackson

211030943

Thiago França

170119429

Preparação da Solução para o Lançamento

Integração com Sistemas Existentes

Para prever o vencedor das partidas da NBA utilizando dois algoritmos baseados nos modelos de aprendizado de máquina Random Forest e o outro Regressão logística, para esses dois modelos é crucial integrar a solução com os sistemas já existentes. É necessário mapearmos todos os sistemas e pontos de integração, como bancos de dados que contém estatísticas das últimas temporadas, APIs de coleta de dados em tempo real e plataformas de análise de desempenho dos jogadores. Após o mapeamento, devemos desenvolver APIs RESTful seguras e escaláveis que permitam a comunicação entre o modelo e os sistemas de dados existentes. Em seguida estabeleceremos pipelines de dados automatizados para fornecer continuamente dados atualizados ao modelo, utilizando ferramentas de ETL (Extract, Transform, Load) para integrar e preparar os dados. Por fim, implementar os pipelines de CI/CD com ferramentas como Github Actions para automatizar a implantação do modelo preditivo e suas atualizações.

Testes de Estabilidade e Segurança

Antes de prever o vencedor das partidas da NBA, é necessário realizar testes de estresse no sistema para verificar o comportamento do sistema sob cargas intensas e identificar gargalos de desempenho, buscando encontrar e mitigar vulnerabilidades. Com relação a segurança dos dados utilizados para treinamento e teste dos modelos, não seria necessário um cuidado tão rigoroso em relação a dados sensíveis, visto que os dados utilizados são de caráter público, sendo publicados diretamente no próprio site da NBA. Também é importante realizar testes de integração a fim de garantir que todas as partes funcionem corretamente quando trabalhando



juntas, especialmente os módulos de comunicação entre o modelo e os bancos de dados. Com isso realizaremos os testes de regressão visando assegurar que futuras atualizações no modelo não afetem funcionalidades existentes e que a precisão das previsões seja mantida.

Preparação de um Ambiente de Monitoramento

Configurar um ambiente de monitoramento robusto é essencial para acompanhar o desempenho do sistema preditivo em tempo real, para isso, podemos utilizar ferramentas como Prometheus, Grafana ou Kibana para acompanhar métricas do modelo e da infraestrutura. Também é importante implementar um sistema de logging bem detalhado e profundo, buscando obter logs que nos ajudem a compreender e identificar problemas e falhas. Com isso, também podemos gerar relatórios periódicos que nos permitirão analisar o desempenho do modelo e da infraestrutura, permitindo ajustes proativos.

Estratégias de Manutenção e Atualização Contínua

Para garantir uma boa manutenção e atualização do sistema, é de suma importância fazer o re-treinamento dos modelos para manter sua eficiência e evitar que caiam em desuso. Os re-treinamentos podem ser feitos após cada rodada do campeonato da NBA com dados novos dos jogos que aconteceram na rodada e automatizando o pipeline de re-treinamento. Tendo em vista que separamos nossos dados por temporadas e em tabelas para temporada regular e playoffs, poderíamos também fazer o re-treinamento duas vezes por ano. Onde os modelos seriam treinados após a temporada regular e depois dos playoffs.

Outros passos seriam implementar um sistema de avaliação contínua para monitorar o desempenho do modelo em tempo real, definindo KPIs (Key Performance Indicators) para medir a eficácia. Estabelecer processos contínuos de coleta e limpeza de dados, garantindo armazenamento seguro e conformidade com regulamentos de privacidade. Realizar backups regulares dos modelos e dados, e estabelecer procedimentos de recuperação para restaurar rapidamente o serviço em caso de falhas.