











Decisões  <p>Como as previsões são usadas para tomar decisões que fornecem o valor proposto ao usuário final?</p> <p>O usuário final são os indivíduos que acessam serviços de streaming de filmes.</p>	Tarefa de ML  <p>Entrada, saída a ser prevista, tipo de problema.</p> <p>Entrada: dataset de avaliação de filmes Saída: pontuação de similaridade, cluster dos itens. Tipo de problema: Clustering, aprendizado estatístico</p>	Proposições de Valor  <p>O que estamos tentando fazer para o(s) usuário(s) final(is) do sistema preditivo? Quais objetivos estamos cumprindo?</p> <p>Os sistemas de recomendação de filmes prevêem quais filmes um usuário gostará com base nos atributos presentes em filmes preferidos.</p>	Fontes de Dados  <p>Quais fontes de dados brutos podemos usar (interno e externo)?</p> <p>Dataset MovieLens</p> <p>TMDB Movie Dataset</p>	Coleta de Dados  <p>Como obtemos novos dados para aprendizado (entradas e saídas)?</p> <p>De datasets públicos do Kaggle.</p>
Execução das Previsões  <p>Quando fazemos previsões sobre novas entradas? Em batch ou online? Quanto tempo temos para trabalhar uma nova entrada e fazer uma previsão?</p> <p>Novas previsões serão executadas quando solicitadas pelos usuários de forma online.</p>	Avaliação Offline  <p>Métodos e métricas para avaliar o sistema antes da implantação.</p> <p>RMSE (Root Mean Square Error) MAE (Mean Absolute Error)</p>	Avaliação e Monitoramento Online  <p>Métodos e métricas para avaliar o sistema após a implantação e para quantificar a criação de valor.</p> <p>Matriz TF-IDF Matriz de gêneros Matriz de pontuações</p>	Atributos (Features)  <p>Representações de entrada extraídas de fontes de dados brutas.</p> <p>elenco, equipe, keywords, budget, gênero, língua, título, sinopse, rating</p>	Criação de Modelos  <p>Quando criamos/atualizamos modelos com novos dados de treinamento? Quanto tempo temos para trabalhar as entradas e criar/atualizar um modelo?</p> <p>O modelo será atualizado conforme a demanda do usuário.</p>



Decisões

Como as previsões são usadas para tomar decisões que fornecem o valor proposto ao usuário final?

O usuário final são os indivíduos que acessam serviços de streaming de filmes. As previsões serão utilizadas para ajudar aos usuários a escolherem quais filmes eles podem assistir, com base nas preferências deles (gênero, diretor, ator, imdb...)



Tarefa de ML

Entrada, saída a ser prevista, tipo de problema.

Entrada: dataset de avaliação de filmes (título | overview | popularidade | vote_average ...)

Saída: pontuação de similaridade, cluster dos itens.

Tipo de problema: Clustering, aprendizado estatístico



Proposições de Valor

O que estamos tentando fazer para o(s) usuário(s) final(is) do sistema preditivo? Quais objetivos estamos cumprindo?

Os sistemas de recomendação de filmes geralmente prevêm quais filmes um usuário gostará com base nos atributos presentes em filmes preferidos anteriormente. Esses sistemas de recomendação são benéficos para organizações que coletam dados de grandes quantidades de clientes e desejam fornecer efetivamente as melhores sugestões possíveis.



Fontes de Dados

Quais fontes de dados brutos podemos usar (interno e externo)?

Dataset MovieLens

(<https://www.kaggle.com/rounakbani-k/the-movies-dataset>)

TMDB Movie Dataset

(<https://www.kaggle.com/tmdb/tmdb-movie-metadata>)



Coleta de Dados

Como obtemos novos dados para aprendizado (entradas e saídas)?

De datasets públicos do Kaggle.



Execução das Previsões

Quando fazemos previsões sobre novas entradas? Em batch ou online? Quanto tempo temos para trabalhar uma nova entrada e fazer uma previsão?

Novas previsões serão executadas quando solicitadas pelos usuários de forma online.

Tempo de previsão: 8h



Avaliação Offline

Métodos e métricas para avaliar o sistema antes da implantação.

RMSE (Root Mean Square Error)
MAE (Mean Absolute Error)



Avaliação e Monitoramento Online

Métodos e métricas para avaliar o sistema após a implantação e para quantificar a criação de valor.

Matriz TF-IDF
Matriz de gêneros
Matriz de pontuações



Atributos (Features)

Representações de entrada extraídas de fontes de dados brutas.

- **elenco**
- **equipe**
- **keywords**
- **budget**
- **gênero**
- **língua**
- **título**
- **sinopse**
- **rating**



Criação de Modelos

Quando criamos/atualizamos modelos com novos dados de treinamento? Quanto tempo temos para trabalhar as entradas e criar/atualizar um modelo?

O modelo será atualizado conforme a demanda do usuário. Estimamos um tempo de 3 horas para atualização do modelo.