## **Alunos: Felipe Beppler, Gustavo Schneider Rodrigues**

## **a. Problema:**

O problema escolhido é a recomendação de filmes baseada em regras de associação. O objetivo é identificar padrões de filmes que são frequentemente assistidos juntos para criar recomendações personalizadas aos usuários, similar ao que plataformas de streaming utilizam ("quem assistiu X também assistiu Y").

## **b. Dados utilizados para modelagem do problema (dataset):**

Utilizei o dataset MovieLens real (versão small - 100K), que contém:

* Avaliações de filmes dadas por usuários (escala de 0.5 a 5.0 estrelas)
* Informações sobre filmes, incluindo títulos e gêneros
* O dataset é composto por milhares de avaliações feitas por usuários sobre milhares de filmes

## **c. Passos utilizados para geração de regras:**

1. **Obtenção e preparação dos dados**:
   * Download do dataset MovieLens 100K da fonte oficial
   * Carregamento dos arquivos de avaliações (ratings.csv) e filmes (movies.csv)
   * Filtragem para considerar apenas filmes "assistidos" (avaliação ≥ 3.5)
2. **Pré-processamento e transformação**:
   * Foco nos 500 filmes mais populares para gerenciar o uso de memória
   * Seleção de usuários ativos (que avaliaram pelo menos 5 filmes)
   * Transformação dos dados em formato de transações, onde cada transação representa o conjunto de filmes assistidos por um usuário
   * Conversão de IDs de filme para títulos para melhor interpretabilidade
3. **Mineração de padrões frequentes**:
   * Aplicação do algoritmo Apriori com suporte mínimo de 2%
   * Aplicação do algoritmo FP-Growth com o mesmo suporte mínimo
   * Medição do tempo de execução de cada algoritmo para comparação de performance
4. **Geração das regras de associação**:
   * Transformação dos padrões frequentes em regras de associação
   * Utilização de confiança mínima de 30%
   * Cálculo de métricas importantes: suporte, confiança e lift
5. **Análise das regras**:
   * Comparação do número de regras geradas por cada algoritmo
   * Verificação da sobreposição entre as regras dos dois algoritmos
   * Análise estatística das métricas de qualidade das regras
   * Identificação das regras com maior lift (mais relevantes)
6. **Implementação do sistema de recomendação**:
   * Desenvolvimento de uma função que utiliza as regras para recomendar novos filmes
   * O sistema considera os filmes já assistidos pelo usuário
   * Ranking das recomendações baseado em um score combinado de suporte, confiança e lift

## **d. Regras Geradas:**

As regras geradas pelo sistema têm o formato "Se um usuário assistiu ao filme A (e possivelmente B, C...), então provavelmente também gostará do filme X".

Exemplos de regras geradas (com dados reais do MovieLens):

* "Star Wars: Episode IV - A New Hope (1977)" => "Star Wars: Episode V - The Empire Strikes Back (1980)"
* "The Lord of the Rings: The Fellowship of the Ring (2001)" => "The Lord of the Rings: The Two Towers (2002)"

Cada regra inclui métricas importantes:

* **Suporte**: Frequência com que os filmes aparecem juntos no dataset (percentual de usuários que assistiram ambos os filmes)
* **Confiança**: Probabilidade condicional de assistir ao filme consequente dado que assistiu ao antecedente
* **Lift**: Razão entre a confiança observada e a confiança esperada se os filmes fossem independentes (valores acima de 1 indicam correlação positiva)

A comparação entre Apriori e FP-Growth mostra que o segundo algoritmo é mais eficiente em termos de tempo de execução (especialmente em datasets grandes), enquanto gera o mesmo conjunto de regras. Isso confirma a vantagem de usar FP-Growth para datasets volumosos como o MovieLens completo.

O sistema de recomendação desenvolvido utiliza estas regras para sugerir novos filmes aos usuários com base em seu histórico de visualização, priorizando recomendações com maior score combinado (produto das métricas de suporte, confiança e lift).