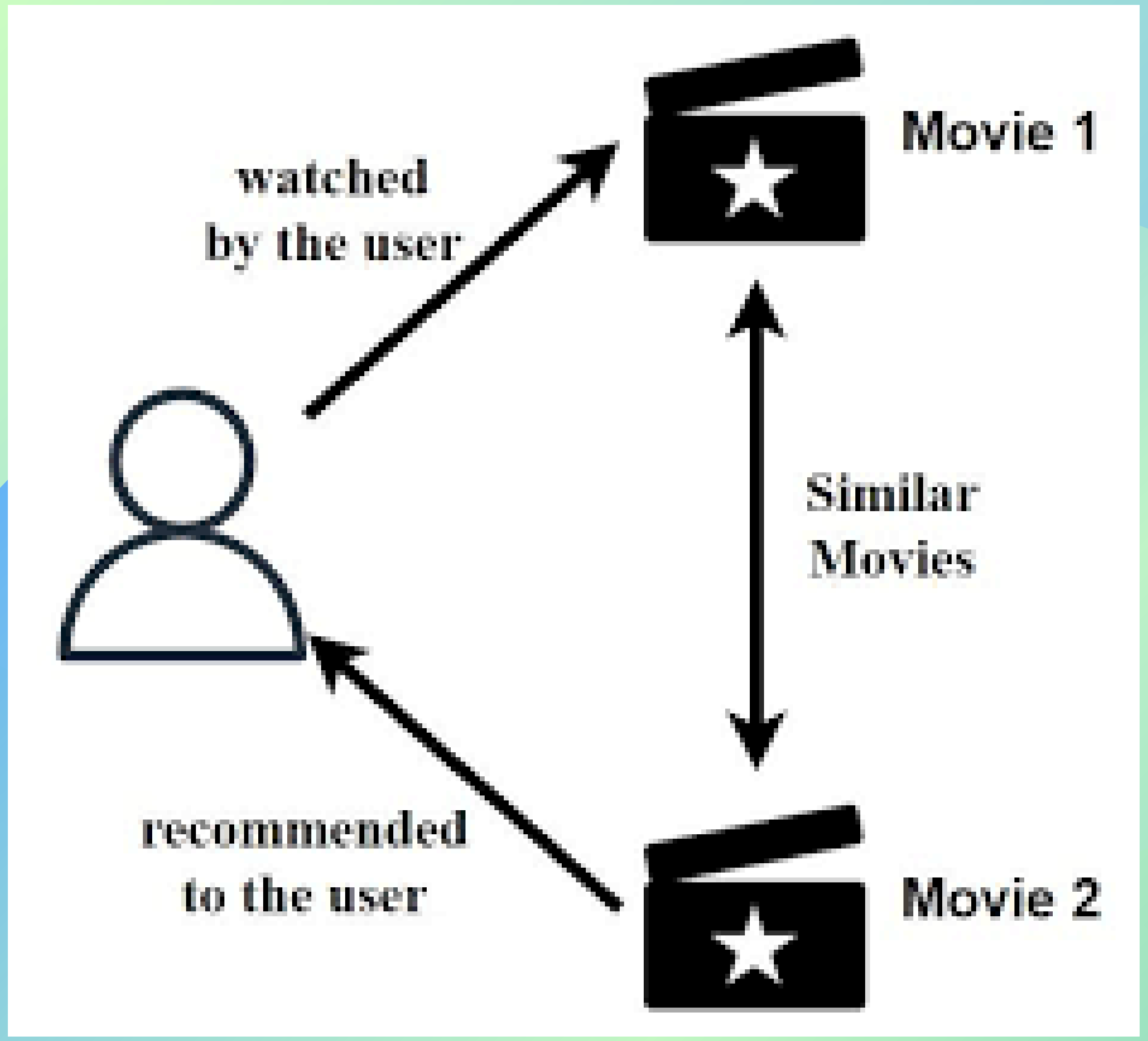


This report comprises the following explanations:

Justification of the selection of algorithm

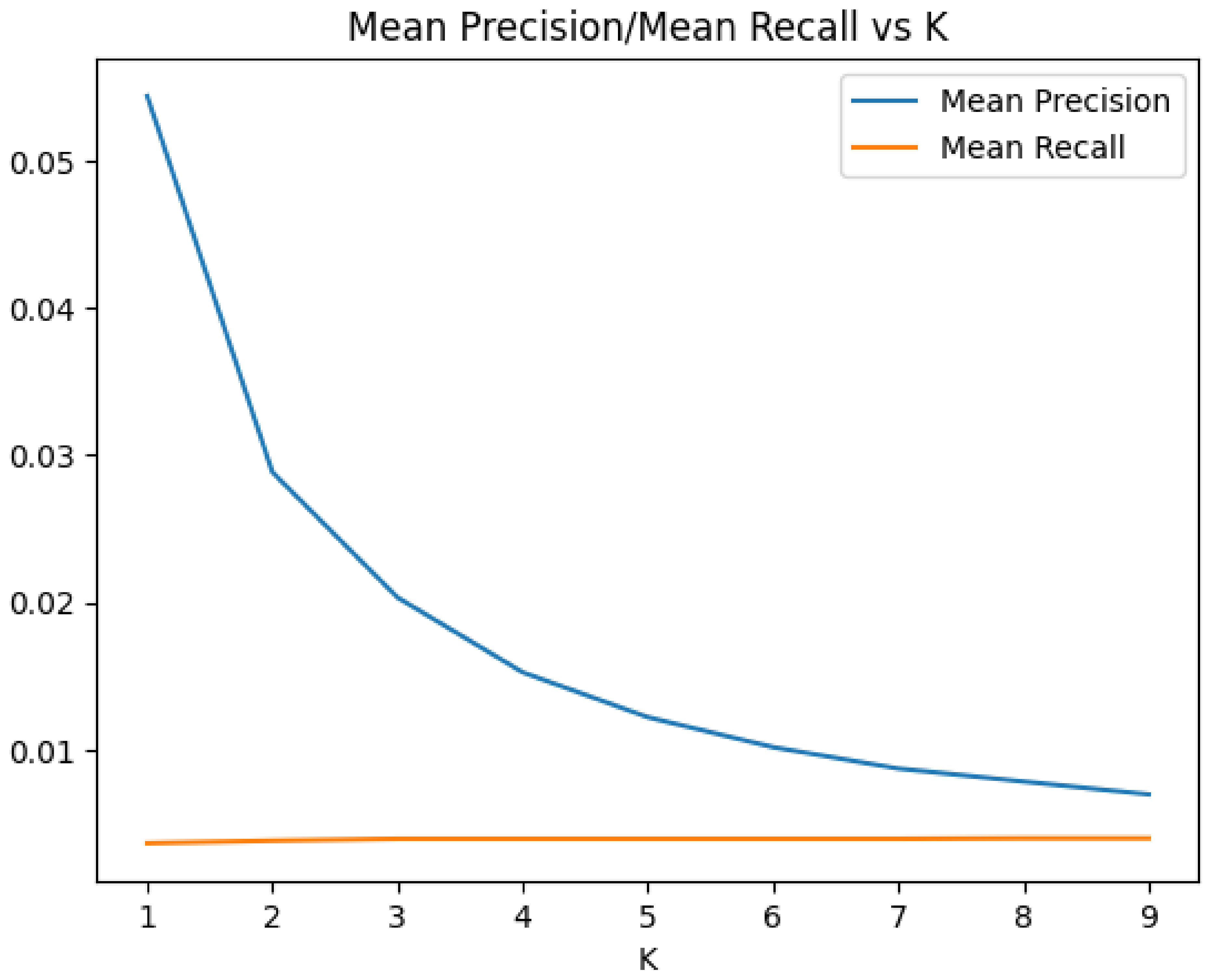
A brief explanation of how the entire recommendation system is built

Team →

Varghese P Kuruvilla
varghese.kuruvilla@research.iiit.ac.in
Roll No: 2023701007

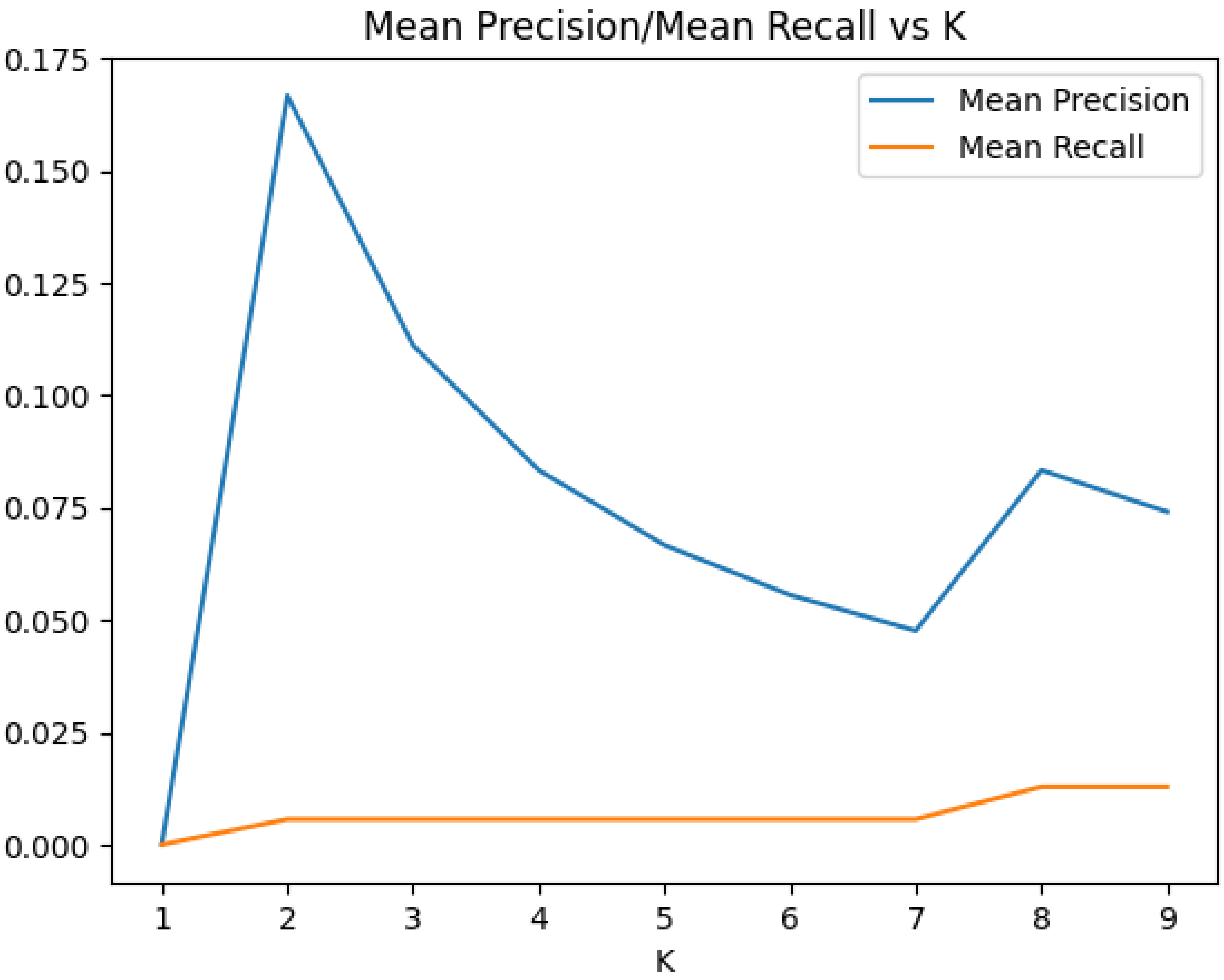
Deepti Rawat
deepti.rawat@research.iiit.ac.in
Roll No: 2022801016

Plots →



- Graphs of the average precision and average recall values are also plotted for all the users while varying k (the number of rules) from 1 to 10.

Plots →



- Graphs of the average precision and average recall values are also plotted for a sample of the users while varying k (the number of rules) from 1 to 10.

Justification of the selection of algorithm

Algorithm Utilized for mining association rules

- FP Growth Algorithm

Reasons for using FP Growth Algorithm

The selection of the FP-growth algorithm for mining association rules in this scenario is justified by several key considerations:

High-Dimensional Item Space:
Given that the training data set consists of entries in the form <user id, {movies rated}>, it implies a potentially large set of movies. The Apriori algorithm, which generates candidate itemsets explicitly, might be computationally expensive due to the vast itemset space. FP-growth is ideal for dealing with large item sets since it does not require generating candidates explicitly, unlike Apriori. This reduces computational complexity and makes it more efficient.

Justification of the selection of algorithm

Reasons for using FP Growth Algorithm

Efficiency in Large Datasets:
Adding to the last point, FP-growth creates a compact data structure called an FP-tree, allowing it to quickly find common itemsets even in massive databases. In contrast, Apriori requires numerous passes through the data, which takes longer and consumes more resources.

Reduced Passes Over Data:
FP-growth constructs a compact data structure and performs fewer passes over the data compared to Apriori, which is helpful when resources are limited. It also speeds up rule mining.

A brief explanation of how the entire recommendation system is built

Steps followed to build Recommendation System

- We've leveraged a data mining approach: FP-growth algorithm, to construct a movie recommendation system that provides personalized movie recommendations based on association rules derived from user behavior. Here is a concise explanation of the steps:

Data Preprocessing

- First, we acquire a training dataset consisting of entries in the format <user id, {movies rated above 2}>. This dataset serves as the foundation for mining association rules.

Association Rule Mining

- Then, we utilize the FP-growth algorithm, a popular method for mining frequent itemsets in large datasets, to extract association rules of the form X→Y.
- Specifically, we focus on rules where X contains a single movie and Y comprises sets of movies from the training data.

A brief explanation of how the entire recommendation system is built

Rule Generation

- Extracted association rules X→Y represent patterns of co-occurring movies. These rules imply that if a user has rated movie X, they are likely to be interested in movies in set Y.

Confidence and Support Thresholds

- Set appropriate confidence and support thresholds based on domain knowledge and data characteristics. These thresholds determine the strength and relevance of the discovered rules. In our case, we have set min_sup as 50 and min_conf as 0.1

Recommendation Generation

- Given a user's interaction (e.g., movies they have rated), we can apply the extracted association rules to generate personalized recommendations.
- Identify movies rated by the user as X, then find associated sets Y from the rules. These Y sets represent potential recommendations.

Thank You