**Title:** Enhancing Book Recommendation Systems through Associative Data Mining

**Group Members:** Dev Gaur (2020B5A72178H), P Sai Aditya (2020B5A70987H)

**Data Source:** Our project will leverage a comprehensive dataset of Amazon book ratings by users. This dataset includes Online data for books from Amazon along with user ratings and users who bought them.

We would be using 2 datasets, one having the book details(like BookId, Name, Author, etc), other consisting of user and their ratings of different books they have bought and read from Amazon.

**Problem Statement:** Finding relevant books in the vast sea of available literature can overwhelm readers. Our project aims to solve the challenge of improving book recommendation systems by implementing associative data mining techniques. By analyzing patterns in book ratings on Amazon, we intend to uncover relationships between books that users rate similarly. This will enable us to recommend books a user will likely enjoy, based on their rating history and the discovered associations.

**Significance of the Problem:** The problem is intriguing because it intersects user experience and technology. Enhancing recommendation systems can significantly impact readers' satisfaction and engagement, helping them discover books that match their preferences and reading habits. Moreover, understanding complex patterns in book ratings can provide insights into broader trends in reading behaviours and preferences.

**Innovation and Learning Objectives:** Our project is unique in its application of associative data mining techniques to the domain of book recommendations. Each team member will implement a distinct algorithm, such as Apriori, FP-growth, or Eclat, to uncover associations between books. This approach will improve our understanding of data mining algorithms and their practical applications and allow us to compare the effectiveness of different techniques in real-world scenarios. Through this project, we aim to learn about the intricacies of associative data mining, its challenges, and its potential to enhance digital platforms and services.

We want to test how much better the FP tree algorithm is to the Apriori algorithm in the sense of time consumed and is it also better in terms of space.

Similar testing would be done between the ECLAT algorithm and the aforementioned algorithms.

We would also understand how using the ECLAT algorithm make certain calculations easier to compute and would present additional data observations that we can obtain easily.

We would like to make various observations such as:

1. Finding the similar set of books and how many of them belong to the same author/genre.
2. Are 2 or more genre books bought together i.e., is there a match between genres.
3. How frequent is a book, if it is in multiple frequent item sets. It must be a famous one and so on similar observations will be taken.