

# DATA-643 Final Project Plan

*Mohamed Elmoudni*

*Shazia Khan*

*Senthil Dhanapal*

## Contents

Introduction . . . . .	1
Purpose . . . . .	1
Data acquisition . . . . .	1
Data Exploration and Preparation . . . . .	2
Model creation . . . . .	2
Model Selection . . . . .	2
Prediction . . . . .	2
Model tuning . . . . .	2

## Introduction

Our project will be about recommending books based on user rating. The dataset is called a Book-Crossings. it is a book ratings dataset compiled by Cai-Nicolas Ziegler based on data from bookcrossing.com. It contains 1.1 million ratings of 270,000 books by 90,000 users. The ratings are on a scale from 1 to 10, and implicit ratings are also included.

Our project plan will be based on the below high level steps:

- 1- Data acquisition
- 2- Data Exploration and Preparation
- 3- Model creation
- 4- Model Selection
- 5- Prediction
- 6- Model tuning

## Purpose

The goal of the project is to build a comprehensive recommendation system for books based on user ratings. We also take into account user's location and age to improve relevance of recommendation.

Below is detailed highlight for the each section in our project plan:

## Data acquisition

- a. Data source Identification
- b. Data collection and storing
- c. Data merging

## Data Exploration and Preparation

- a. Variable Identification
- b. Univariate Analysis
- c. Bi-variate Analysis
- d. Missing values treatment
- e. Outlier treatment

## Model creation

- a. Model identification
- b. Model building
- c. Model creation

## Model Selection

- a. Model comparison
- b. Model selection

## Prediction

- a. Test data
- b. Prediction on test/unseen data

## Model tuning

- a. Test the model using data augmentation
- b. Tune the model using different functions