



Building Recommendation Engine on Books datasets

Team members :

1. Raja Sekhar Pothina

2. Chandrashekar Akkenapally



Concept Description

To understand and implement a simple recommendation system using various Data Mining techniques.

The main aim of the project is to classify the data extracted from csv files into different categories of reader's choice.

The techniques used in the project are K-Nearest Neighbors, Kmeans, 1R Classifier and Cosin Similarity.



Data Collection

The data is provided by the professor.

The provided data set consists of three csv files, one is the book_tags.csv, second is tags.csv and third is goodreads_books.csv file.



Handling Duplicate and Missing Values

Handling a row having same title but different tag-names to have tag-names which has max count.

Books without title are removed from the dataframe by merging.



Data Pre-Processing

Initializing the Series,
Coverlink columns
with Default values

Combining series and
author name with title
name to form a single
column to recognize
each book differently.



Data Exploration

To generate the overall statistics of the books data.

From the books data we try to find the most common books in the list.

We will also try get the list of top-rated books from the data set.

Mining for the **authors publishing most of the books.**

Data Exploration

Finding out the best Author as the author who has written books having best average rating ≥ 4.3 and rating count > 50000 .

Calculating the rating distribution for the books .

Checking for relation between average_rating and review counts

Mining for the books **with most reviews**.

Data Mining Techniques

K Nearest Neighbors

- Books Recommended for Stardust by Neil Gailman

title: Room, author: Emma Donoghue



title: YÃ¼zyÄ±llÄ±k YalnÄ±zlik, author: Gabriel GarcÄ±a MÃ±rquez, SeÅŸkin Selvi



title: One Hundred Years of Solitude, author: Gabriel GarcÄ±a MÃ±rquez, Gregory Rabassa



title: Cien aÃ±os de soledad, author: Gabriel GarcÄ±a MÃ±rquez



title: Cem Anos de SolidÃ£o, author: Gabriel GarcÄ±a MÃ±rquez, Margarida Santiago



Data Mining Techniques

K Nearest Neighbors

- Books Recommended for Book titled Room

title: Twilight: The Complete Illustrated Movie Companion, series: (The Twilight Saga: The Official Illustrated Movie Companion #1), author: Mark Cotta Vaz



title: I Know Why the Caged Bird Sings, series: (Maya Angelou's Autobiography #1), author: Maya Angelou



title: The Screwtape Letters, author: C.S. Lewis



title: The One, series: (The Selection #3), author: Kiera Cass



title: Safe Haven, author: Nicholas Sparks



Data Mining Techniques

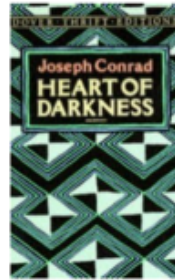
Kmeans Classifier

- The cluster of Stardust books

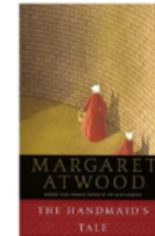
title: Shiver, series: (The Wolves of Mercy Falls #1), author: Maggie Stiefvater



title: Heart of Darkness, author: Joseph Conrad



title: The Handmaid's Tale, series: (The Handmaid's Tale #1), author: Margaret Atwood



title: Diary of a Wimpy Kid, series: (Diary of a Wimpy Kid #1), author: Jeff Kinney



title: Anna and the French Kiss, series: (Anna and the French Kiss #1), author: Stephanie Perkins

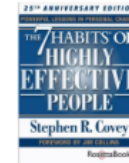


Data Mining Techniques

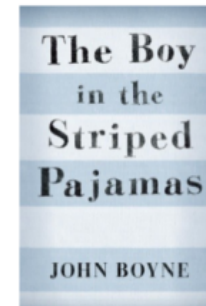
- The cluster of Room books

Kmeans Classifier

title: The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change, author: Stephen R. Covey



title: The Boy in the Striped Pajamas, author: John Boyne



title: Cinder, series: (The Lunar Chronicles #1), author: Marissa Meyer



title: Fifty Shades Darker, series: (Fifty Shades #2), author: E.L. James



title: The Last Olympian, series: (Percy Jackson and the Olympians #5), author: Rick Riordan



Data Mining Techniques

- Recommending based on average rating greater than or equal to the book selected

- Stardust Recommendation

title: This Tender Land, author: William Kent Krueger



title: Shadow of a Human, series: (The Human Cycle #2), author: JD Estrada



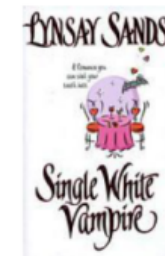
title: The Crossing, series: (The Border Trilogy #2), author: Cormac McCarthy



title: Friction in Motion, author: Sy Kadella



title: Single White Vampire, series: (Argeneau #3), author: Lynsay Sands



1R
Classifier

Data Mining Techniques

- Room Book Recommendation

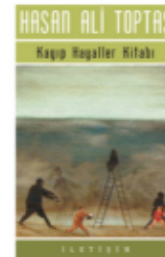
title: The Poor Mouth: A Bad Story about the Hard Life, author: Flann O'Brien, Patrick C. Power, Ralph Steadman



title: The Lightning-Struck Heart, series: (Tales From Verania #1), author: T.J. Klune



title: Kayıp Hayaller Kitabı, author: Hasan Ali Toptaş



title: Le grand cahier, series: (La trilogie des jumeaux #1), author: Agota Kristof



title: Without Fail, series: (Jack Reacher #6), author: Lee Child, Dick Hill



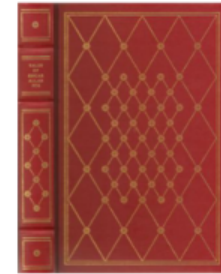
1R
Classifier

Data Mining Techniques

- Recommending Books based on author of the book selected

- Stardust Recommendation

title: Tales of Edgar Allan Poe, author: Edgar Allan Poe, Harry Clarke



title: The Sandman Omnibus, Vol. 1, series: (The Sandman Omnibus #1), author: Neil Gaiman

title: The Sandman Omnibus, Vol. 2, series: (The Sandman Omnibus #2), author: Neil Gaiman



title: A Calendar of Tales, author: Neil Gaiman



title: Graceling, series: (Graceling Realm #1), author: Kristin Cashore



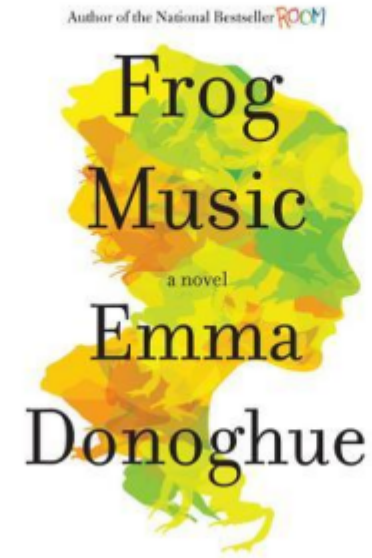
1R
Classifier

Data Mining Techniques

- Room Book Recommendation

**1R
Classifier**

title: Frog Music, author: Emma Donoghue



Data Mining Techniques

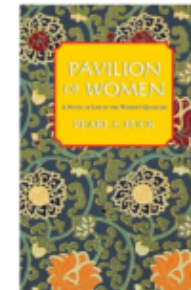
- Recommending Books Based on tag name of the selected Books

- Stardust Recommendation

title: Bad News/Good News, series: (Beacon Street Girls #2), author: Annie Bryant



title: Pavilion of Women, author: Pearl S. Buck



title: Finnikin of the Rock, series: (Lumatere Chronicles #1), author: Melina Marchetta



title: Conspiracy of Silence, series: (Tox Files #1), author: Ronie Kendig



title: The BLDGBLOG Book, author: Geoff Manaugh



1R
Classifier

Data Mining Techniques

- Room Book Recommendation

1R Classifier

title: 13, series: (Tallent & Lowery #1), author: Amy Lignor



title: Memory, series: (Vorkosigan Saga (Publication Order) #10), author: Lois McMaster Bujold



title: Dancing With Werewolves, series: (Delilah Street, Paranormal Investigator #1), author: Carole Nelson Douglas



title: Dreams of Joy, series: (Shanghai Girls #2), author: Lisa See



title: The Destiny of Violet & Luke, series: (The Coincidence #3), author: Jessica Sorensen



Data Mining Techniques

- Recommending Books based on review count of the selected Book

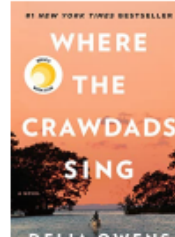
- Stardust Recommendation

1R
Classifier

title: A Thousand Splendid Suns, author: Khaled Hosseini



title: Where the Crawdads Sing, author: Delia Owens



title: The Hitchhiker's Guide to the Galaxy, series: (Hitchhiker's Guide to the Galaxy #1), author: Douglas Adams



title: Beautiful Creatures, series: (Beautiful Creatures #1), author: Kami Garcia, Margaret Stohl



title: Lovely Trigger, series: (Tristan & Danika #3), author: R.K. Lilley



Data Mining Techniques

- Room Book Recommendation

title: Crime and Punishment, author: Fyodor Dostoyevsky, David McDuff, Fyodor Dostoyevsky, Michael R. Katz



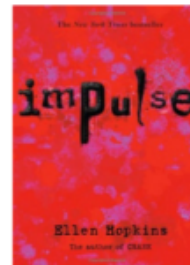
title: Leaves of Grass, author: Walt Whitman



title: City of Bones, series: (The Mortal Instruments #1), author: Cassandra Clare



title: Impulse, series: (Impulse #1), author: Ellen Hopkins



title: Girl with a Pearl Earring, author: Tracy Chevalier



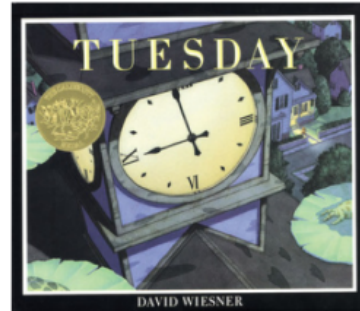
1R
Classifier

Data Mining Techniques

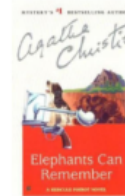
Cosine Similarity

- Stardust recommendations

title: Tuesday, author: David Wiesner



title: Elephants Can Remember, series: (Hercule Poirot #40), author: Agatha Christie



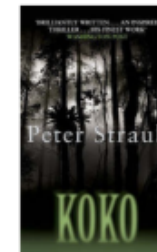
title: Time And Eternity: Exploring God's Relationship To Time, author: William Lane Craig



title: Sashenka, series: (Moscow Trilogy #1), author: Simon Sebag Montefiore



title: Koko, series: (Blue Rose Trilogy #1), author: Peter Straub

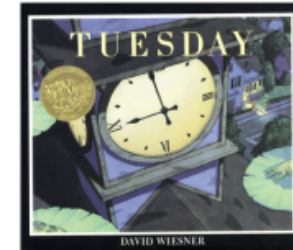


Data Mining Techniques

Cosine Similarity

- Room book recommendations

title: Tuesday, author: David Wiesner



title: Elephants Can Remember, series: (Hercule Poirot #40), author: Agatha Christie



title: Time And Eternity: Exploring God's Relationship To Time, author: William Lane Craig



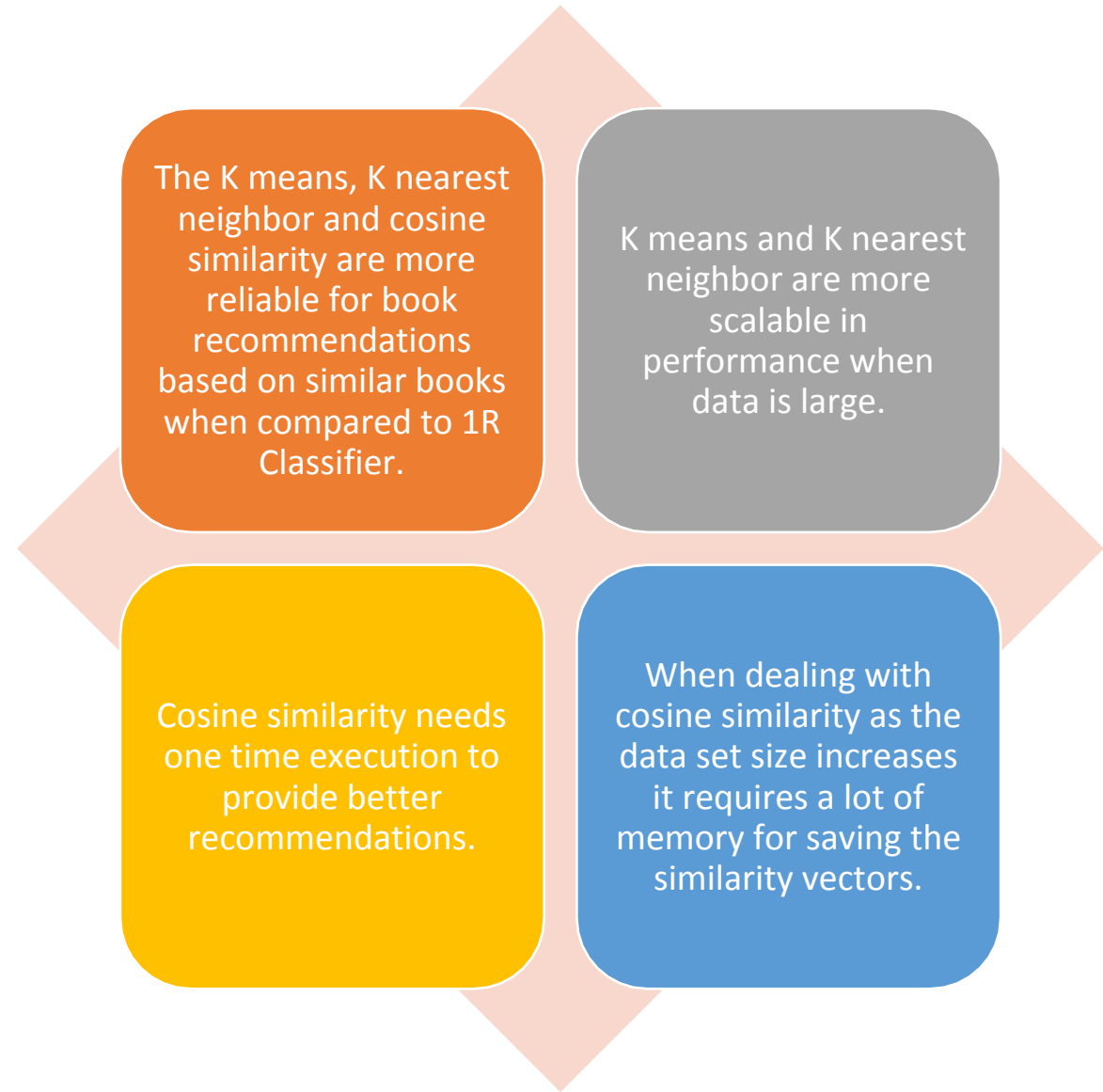
title: Sashenka, series: (Moscow Trilogy #1), author: Simon Sebag Montefiore



title: Koko, series: (Blue Rose Trilogy #1), author: Peter Straub



Results



Reference

- [NumPy.Retrieved \(2022, Jul 27\)](#)
- [Pandas Package. Retrieved \(2022, Jul 27\)](#)
- [Matplotlib.Retrieved \(2022, Jul 27\)](#)
- [Seaborn.Retrieved \(2022, Jul 27\)](#)
- [Sk learn. Retrieved \(2022, Jul 27\)](#)
- [Syntax for filter function \(2022, July 27\)](#)
- [To read specific columns from CSV \(2022, July 27\)](#)
- [Selecting Columns from group_by object in pandas](#)
- [How to change the figure size for displot](#)
- [Recommendation System Algorithms: An Overview](#)

