Data Driven - Assignment 3

Azamat Shora

September 27, 2023

1 Dataset

1.1 Generate list of 15 Popular films by IMDB

First of all I created a separate sheet where I kept 15 films from IMBD site from more popular to less popular according to critics and viewers.

Film Title
The Shawshank Redemption (1994)
The Dark Knight (2008)
Schindler's List (1993)
The Lord of the Rings: The Return of the King (2003)
The Lord of the Rings: The Fellowship of the Ring (2001)
Forrest Gump (1994)
Fight Club (1999)
The Lord of the Rings: The Two Towers (2002)
Inception (2010)
Star Wars: Episode V - The Empire Strikes Back (1980)
The Matrix (1999)
Spider-Man: Across the Spider-Verse
Seven (1995)
Interstellar
Rush

Table 1: List of Popular Films by IMDB version

I also created a separate sheet where I kept personal preferences of films in order from more popular to less popular.

Link to the code : **link**

1.2 Generate 59 random pairs of movies.

In order to generate the 59 pairs, I shuffled the original copied dataset to try to make all the films as equal as possible.

1.3 Annotate each generated pair with your preference vote: the left object is more preferred, or the right object is more preferred.

Next, I created a function that could instead of me choose which film I would like more, as I had already created a separate sheet with the films I prefer and in what order. After

Film Title
Inception (2010)
The Matrix (1999)
Interstellar
The Shawshank Redemption (1994)
The Dark Knight (2008)
Schindler's List (1993)
Fight Club (1999)
Forrest Gump (1994)
The Lord of the Rings: The Fellowship of the Ring (2001)
The Lord of the Rings: The Two Towers (2002)
The Lord of the Rings: The Return of the King (2003)
Star Wars: Episode V - The Empire Strikes Back (1980)
Spider-Man: Across the Spider-Verse
Seven (1995)
Rush

Table 2: List of personally loved Films

that to each pair, I could immediately add either 0 to the first film and 1 to the second film.

For example: [Inception (2010), Seven (1995), 0].

This means that Inception (2010) is more favorable for me personally

2 Tasks

2.1 Collect and annotate movie data

The collected movie data with preferences looks like this

Movie 1	Movie 2	Pref.
Shawshank Redemption (1994)	Inception (2010)	1
Fight Club (1999)	Matrix (1999)	1
LOTR: Return of the King (2003)	Fight Club (1999)	1
Dark Knight (2008)	Forrest Gump (1994)	0
Matrix (1999)	Rush	0

Table 3: Movie Pair Preferences

2.2 Fit the Bradley–Terry model to movie data and report the results

	The Shawshank Redemption (1994)	The Dark Knight (2008)	Schindler's List (1993)	The Lord of the Rings: The Return of the King (2003)	The Lord of the Rings: The Fellowship of the Ring (2001)	Forrest Gump (1994)	Fight Club (1999)	The Lord of the Rings: The Two Towers (2002)	Inception (2010)	Star Wars: Episode V - The Empire Strikes Back (1980)	The Matrix (1999)	Spider-Man: Across the Spider-Verse	Seven (1995)	Interstellar	Rush
The Shawshank Redemption (1994)	NaN	0.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
The Dark Knight (2008)	0.0	NaN	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0
Schindler's List (1993)	0.0	0.0	NaN	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
The Lord of the Rings: The Return of the King (2003)	0.0	0.0	0.0	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
The Lord of the Rings: The Fellowship of the Ring (2001)	0.0	0.0	0.0	0.0	NaN	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0
Forrest Gump (1994)	0.0	0.0	0.0	0.0	1.0	NaN	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0
Fight Club (1999)	0.0	0.0	0.0	1.0	0.0	1.0	NaN	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
The Lord of the Rings: The Two Towers (2002)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	1.0	0.0	1.0	0.0	0.0	1.0
Inception (2010)	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	NaN	0.0	1.0	0.0	0.0	1.0	1.0
Star Wars: Episode V - The Empire Strikes Back (1980)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	1.0	1.0	0.0	1.0
The Matrix (1999)	1.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	NaN	0.0	1.0	0.0	1.0
Spider-Man: Across the Spider-Verse	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	1.0
Seven (1995)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	1.0
Interstellar	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	NaN	1.0
Rush	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN

Figure 1: Wins between teams

This data represents 15 films. Each row is the number of times a film has beaten another one (as indicated by the column). So, for the first row, it has bitten 4th film once. A team cannot win itself, and that is why the diagonal entries are missing.

Movie Title	Score
Inception (2010)	8.722×10^{-1}
Interstellar	5.366×10^{-2}
The Matrix (1999)	5.317×10^{-2}
The Dark Knight (2008)	1.335×10^{-2}
The Shawshank Redemption (1994)	3.562×10^{-3}
Schindler's List (1993)	3.490×10^{-3}
Fight Club (1999)	4.264×10^{-4}
Forrest Gump (1994)	7.023×10^{-5}
The Lord of the Rings: The Two Towers (2002)	3.019×10^{-5}
The Lord of the Rings: The Return of the King (2003)	1.045×10^{-5}
The Lord of the Rings: The Fellowship of the Ring (2001)	1.014×10^{-5}
Star Wars: Episode V - The Empire Strikes Back (1980)	9.831×10^{-7}
Seven (1995)	3.143×10^{-8}
Spider-Man: Across the Spider-Verse	2.891×10^{-8}
Rush	0.000

Table 4: Movie Scores

2.3 Calculate Spearman's p between the obtained ranking and IMDb ranking

The calculated Spearman's value: 0.3107142857142856

2.4 Evaluate the results: does the ranking reflect your preferences?

My Preferred Films	Measured Preferred Films
Inception (2010)	Inception (2010)
The Matrix (1999)	The Matrix (1999)
Interstellar	Interstellar
The Shawshank Redemption (1994)	The Shawshank Redemption (1994)
Fight Club (1999)	The Dark Knight (2008)
The Dark Knight (2008)	Schindler's List (1993)
Schindler's List (1993)	Fight Club (1999)
Forrest Gump (1994)	Forrest Gump (1994)
The Lord of the Rings: The Two Towers (2002)	The Lord of the Rings: The Fellowship of the Ring (2001)
The Lord of the Rings: The Fellowship of the Ring (2001)	The Lord of the Rings: The Two Towers (2002)
Star Wars: Episode V - The Empire Strikes Back (1980)	The Lord of the Rings: The Return of the King (2003)
The Lord of the Rings: The Return of the King (2003)	Star Wars: Episode V - The Empire Strikes Back (1980)
Spider-Man: Across the Spider-Verse	Spider-Man: Across the Spider-Verse
Seven (1995)	Seven (1995)
Rush	Rush

Table 5: Movie Pairs comparison

As we can see from the left and right columns it is clear that the overall order of the films is almost the same, and the difference of the indices from the first and second columns for the same film is 1 or 2.

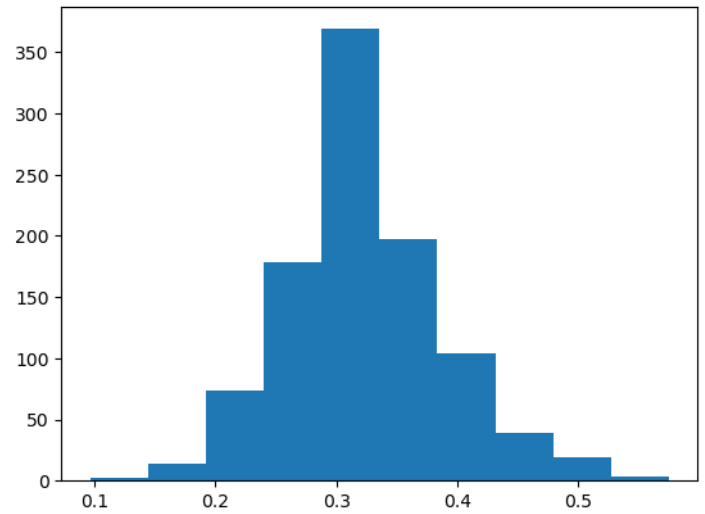## 2.5 Compute bootstrap confidence intervals on your scores

Figure 2: Distribution of Classification Accuracy Using the Bootstrap

Finally, the confidence intervals are reported, showing that there is a 95% likelihood that the confidence interval 20.3571% and 47.5089% covers the true skill of the model.