CAPSTONE PROJECT

FANDANGO MOVIE RATINGS ANALYSIS

Presented By:

1. Student name – k. Kathirvelpandian
2. College name - JP College of Engineering
3. Department - mech
4. Reg no - 951221114501

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

- In recent years, online movie ratings have become a significant influence on audience's viewing choices and the financial success of films. Fandango, a popular online movie ticketing platform, utilizes a five-star rating system that has come under scrutiny. This project aims to analyze Fandango's current movie rating data to determine if there has been any change in their rating system after these allegations. The project will also explore the implications of such a rating system on consumer trust and behavior, as well as the overall impact on the movie industry.

PROPOSED SOLUTION

1. **Perform exploratory data analysis to understand the distribution of ratings.**

2. **Use statistical tests to identify any significant deviations from a normal distribution, which could indicate bias.**

3. **Reverse-engineer the rating algorithm, if possible, to understand how the ratings are calculated.**

4. **Compare the actual user-submitted ratings with the displayed ratings to check for rounding up practices.**

5. **Implement machine learning models to predict unbiased ratings based on user reviews and other metadata.Apply fairness-aware algorithms that can detect and correct for biases.**

   This solution aims to ensure that Fandango's movie ratings are fair, accurate, and trustworthy, ultimately leading to a more reliable platform for both consumers and filmmakers

SYSTEM APPROACH

Building the proposed solution would involve a combination of data processing, feature engineering, and machine learning.Here are the key system and library requirements:

System requirement:

1.Hardware:

- A computer with sufficient processing power, preferably with multiple cores or a GPU for a faster training of machine learning models.
- Adequate RAM to handle the size of the dataset and computational requirements.

2.Software:

- An operating system compatible with the required machine learning libraries(eg.Windows, Linux, maacOS).

Library Requirements:

1.Data Processing and Analysis:

- Pandas: For data manipulation and analysis.
- Numpy: For numerical operations on data.

SYSTEM APPROACH - CONT

2.Data Visualization:

- Matplotlib and Seaborn: For creating visualizations to understand patterns.
- Plotly or Bokeh: Interactive visualization libraries for more complex visualization.

ALGORITHM & DEPLOYMENT

Algorithm Selection

Data Exploration:

- Explore the fandango movie rating dataset's structure, features, and target variable(s).
- Identify potential patterns, correlations, and outliers

Problem Formulation:

- Define the problem:Predicting the consumers in making informed choices about which movies to watch by providing ratings and reviews.

Algorithm Selection

- Regression tasks (eg. Predicting daily rates)

i. consider linear regression, decision trees, or ensemble methods (XGBoost, LightGBM).

- Classification tasks (eg, predicting special tasks)

i. Consider logistic regression,decision trees, or random forests.

ALGORITHM & DEPLOYMENT

Data Input

Data Collection:

Collect a comprehensive dataset of movie ratings from Fandango's website.

Data Cleaning:

Clean the data to remove any duplicates or inconsistencies. Normalize the data to ensure comparability across different movies and time periods.

Feature Engineering:

- Create new features or modify existing ones based on domain knowledge.

ALGORITHM & DEPLOYMENT

Training Process

Data Splitting:

Divide the dataset into training and testing sets to evaluate the model's performance.

Feature Scaling:

Standardize or normalize numerical features to ensure they have a consistent scale.

Model Training:

- Use the selected algorithm to train the model on the training dataset.
- Adjust the hyperparameters to optimize model performance.

Model Evaluation:

Evaluate the model on the testing dataset using appropriate metics (eg, Mean Squared Error for regression, accuracy, precision, recall for classification).

ALGORITHM & DEPLOYMENT

Prediction Process

New Data Input:

Collect new data or use existing data to make predictions.

Preprocessing:

Apply the same data preprocessing steps to the new data.

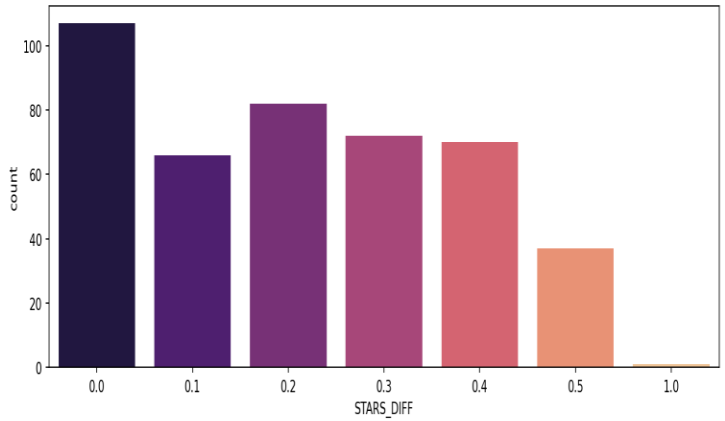
Model Inference:

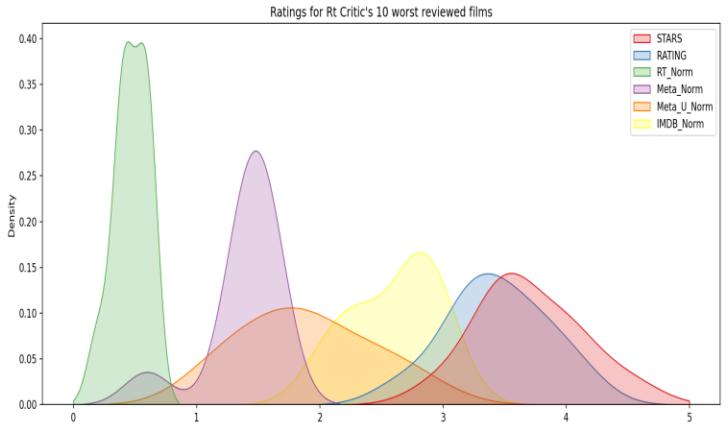
Use the trained model to make predictions on the new data.

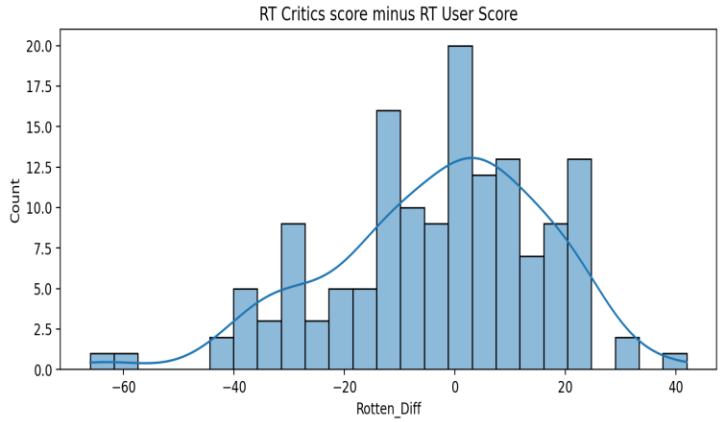
Results Interpretation:

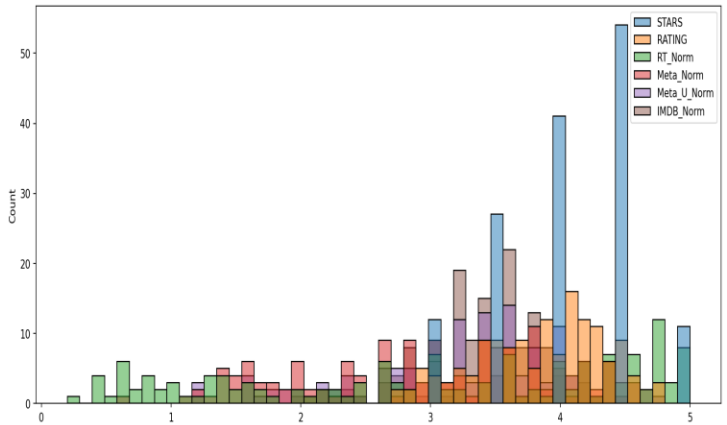
Interpret the model's predictions in the context of the problem at hand.

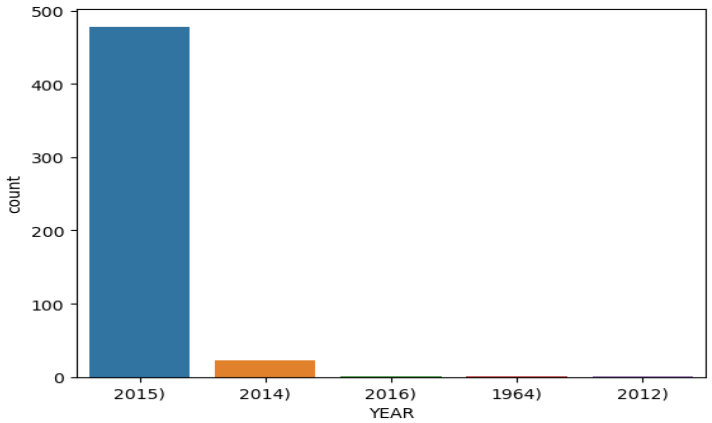
RESULT











edunet foundation

CONCLUSION

- In conclusion, our proposed solution harnesses the power of advanced machine learning algorithms to transform the movie rating process into a dynamic and optimized experience. By metriculously analysing extensive historical movie rating data, we unlock patterns and correlations that are pivotal in addressing key challenges faced by people.

FUTURE SCOPE

The proposed solution lays the foundation for ongoing advancements in the realm of movie rating optimization. Here are key areas for future exploration and enhancement.

Real-time Prediction:

Move towards real-time predictive models that account for instant changes in demand, external events, and other dynamic factors to provide users with up-to-the-minute insights for movie rating decisions.

Personalization and Customization:

Enhance the predictive models to offer more personalized recommendations by considering individual guest preferences, loyalty history, and user-specific requirements, providing a tailored experience for each reviewer.

REFERENCES

- https://www.kaggle.com/datasets
- https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html
- https://seaborn.pydata.org/
- https://matplotlib.org/stable/contents.html

THANK YOU

edunet foundation