

OPEN ENDED ASSIGNMENT

**BUILD AND EVALUATE A MACHINE LEARNING
MODEL TO DECIDE SUITABILITY OF TV/OTT SERIES
FOR CHILDREN'S ENTERTAINMENT**

Dnyanda Patil – UEC2022306 | Ketaki Patil – UEC2022308
Mitali Patil- UEC2022309 | Pranjal Barhate – UEC2022320



INTRODUCTION

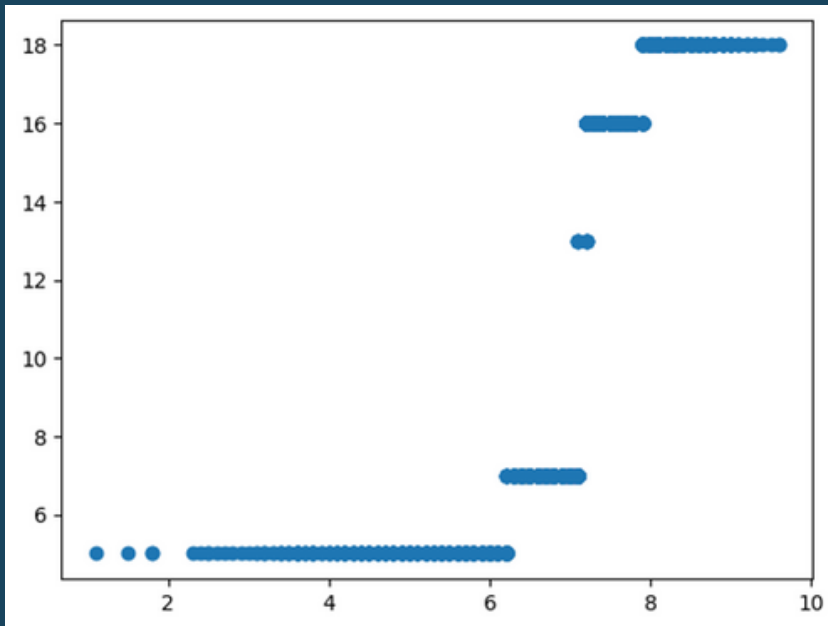
This project aims to use machine learning techniques to assess the appropriateness of TV and OTT series for children. Leveraging algorithms like logistic regression, decision tree, and naive Bayes, we analyze factors such as content ratings and title information to provide insights into suitable content for young audiences. By rigorously evaluating these models against validation datasets, we identify their strengths and limitations, ultimately contributing to safer and more enriching entertainment experiences for children in today's digital age.



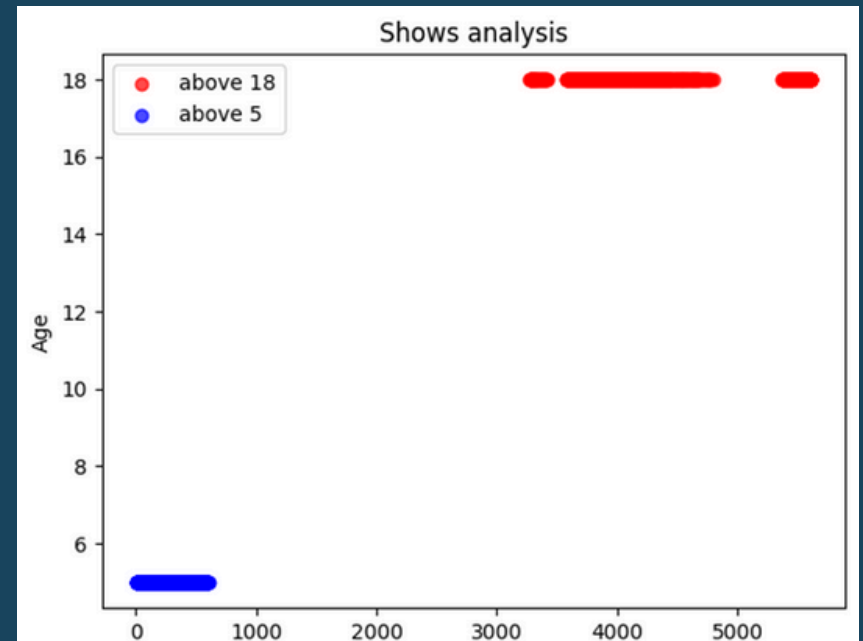
DATA COLLECTION AND PROCESSING

After acquiring the Kaggle dataset, we prioritized data processing. We manually sorted values by attributes like age rating and addressed minimal missing values by removing corresponding rows. These meticulous steps ensured data integrity and prepared the dataset for analysis and model training.

VISUALIZATION



LOGISTIC REGRESSION



NAIVE BAYES

FEATURE SELECTION

We focus on selecting essential features to predict the target variable 'Age'. By selecting these features, we streamline analysis, reduce complexity, and enhance model interpretability. This targeted approach enables more efficient model construction, facilitating the evaluation of TV show suitability for different age groups.

MODEL SELECTION

We implemented three distinct machine learning algorithms: logistic regression, naive Bayes, and decision tree.

After rigorous evaluation and comparison of their performance metrics, it was observed that the decision tree model exhibited the highest accuracy among the three algorithms.

ALGORITHM

- 1. LOAD AND PREPROCESS TV SHOW DATASET, SELECTING RELEVANT FEATURES AND TARGET VARIABLE.**
- 2. SPLIT DATASET INTO TRAINING AND TEST SETS.**
- 3. STANDARDIZE FEATURES FOR CONSISTENT SCALE USING STANDARDSCALER.**
- 4. TRAIN DECISION TREE CLASSIFIER ON TRAINING DATA.**
- 5. EVALUATE MODEL'S PERFORMANCE ON TEST SET, CALCULATING ACCURACY AND CONFUSION MATRIX.**
- 6. FILTER RECOMMENDATIONS FOR AGES 5 TO 16 BASED ON PREDICTED CLASS PROBABILITIES.**
- 7. DISPLAY RECOMMENDED TV SHOW IDS FOR AGES 5 TO 16.**

TESTING AND EVALUATION OF MODEL

Model	Accuracy	Precision	Recall	F1 - Score
Decision Tree	0.996	0.98	0.98	0.98
Naive Bayes	0.990	0.98	0.86	0.90
Logistic Regression	0.977	-	-	-

CONFUSION MATRIX

True\Predicted	Age 5	Age 7	Age 13	Age 16	Age 18
Age 5	170	1	0	0	0
Age 7	0	256	0	0	0
Age 13	0	0	2	0	0
Age 16	0	0	1	297	0
Age 18	0	0	0	1	235

CONCLUSION

In analyzing TV and OTT series suitability for children, decision tree emerged as the top performer with remarkable 99.69% accuracy. Naive Bayes showed respectable 97.0% accuracy, and logistic regression achieved around 97.72%. Decision tree proved most effective, providing superior predictive performance across age categories. Further optimization of naive Bayes and logistic regression models could enhance performance. Employing multiple models and thorough evaluation techniques is crucial for improving content selection for young audiences.



THANK YOU

