MainGames Paidstar Prediction

Overview

New Tab

At MainGames we are committed to help streamers increase their revenues by leveraging the rich source of streaming data to provide actionable insights.

New Tab

In this problem, we want to learn whether certain features of streamers will make them more or less likely to receive stars from their audience, in this case PaidStarPerWatchedHour is our dependent variable.

Outline

EDA

Objective: Overview, Insight and Visualisation Dataset

Investigation	Insight
Structure Data	4250 rows and 127 columns 120 Numeric Feature - 6 categorical Columns, ”PaidStarPerWatchedHour" is a target for prediction.

```
df_train = pd.read_csv('Data Train.csv')
df_train
```

✓ 0.7s

	state	account_length	area_code	international_plan	voice_mail_plan
0	OH	107	area_code_415	no	yes
1	NJ	137	area_code_415	no	no
2	OH	84	area_code_408	yes	no
3	OK	75	area_code_415	yes	no
4	MA	121	area_code_510	no	yes
...	...	...	...	...	...
4245	MT	83	area_code_415	no	no
4246	WV	73	area_code_408	no	no
4247	NC	75	area_code_408	no	no
4248	HI	50	area_code_408	no	yes
4249	VT	86	area_code_415	no	yes

4250 rows × 20 columns

EDA

Investigation	Insight
Quality Data	Tidak ada Duplicate Terdapat beberapa Missing Value Terdapat Inconsistance pada data

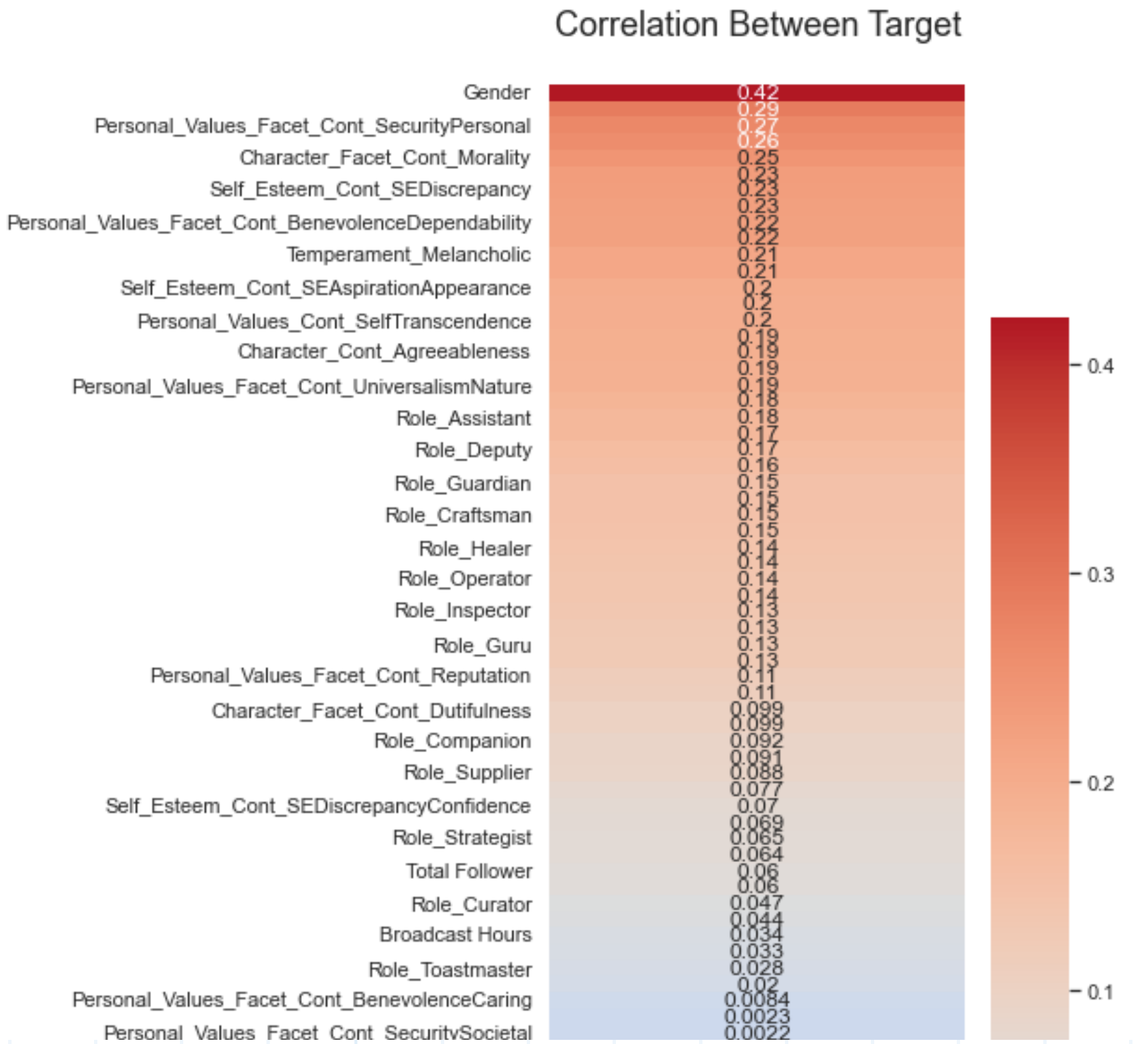
```
[39]  print(df.isna().sum())

...   Country                                  0
      Gender                                   1
      Game                                     2
      Total Follower                           1
      Broadcast Hours                          0
                                              ..
      Character_Facet_Cont_Cautiousness        0
      Role_Curator                             0
      Personal_Values_Cont_Conservation        0
      Self_Esteem_Cont_SELevelIntelligence     0
      Temperament_Centric                      0
      Length: 126, dtype: int64
```

```
Gender
['Male' 'Female' '–' nan]
```

EDA

Objective: Overview, Insight and Visualisation Dataset

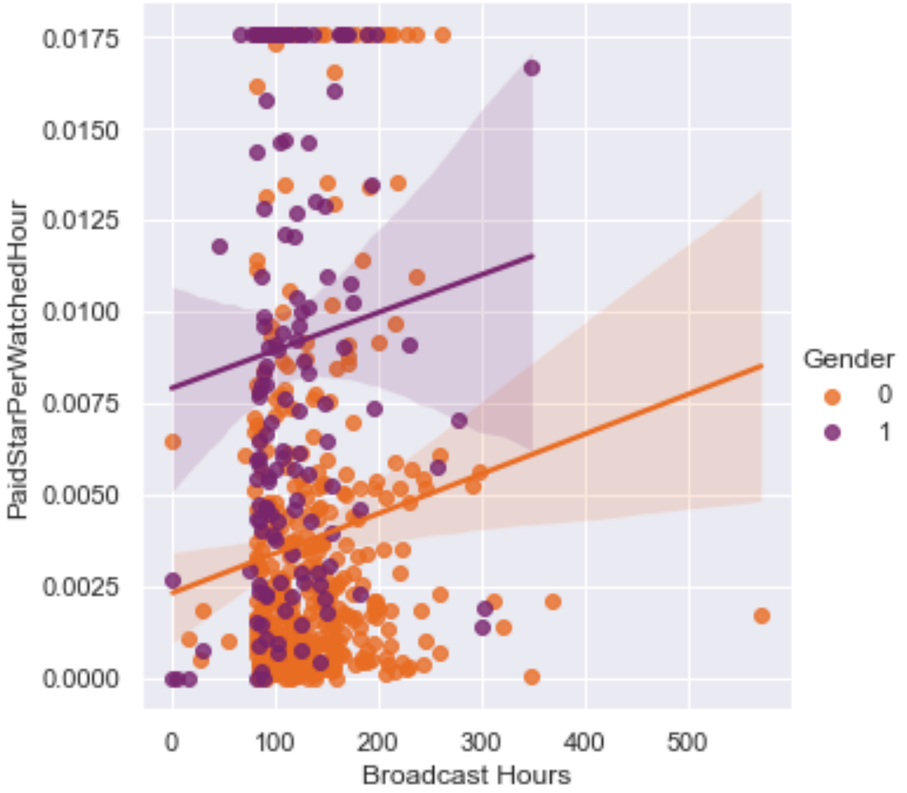
Investigation	Insight
Konten Corelation each feature with the Target	Gender, Personal_Values_Facet_Cont_Security_Personal, Character_Facet_Cont_Morality, Self_Esteem_Cont_SEDiscrepancy, The four mentioned features have a strong correlation with the potential to determine PaidStarPerHour for Streamers. However, it does not exclude the possibility that other features might also have a significant influence on PaidStarPerHour. Therefore, we use feature scoring to select highly relevant features for utilization.



EDA

Objective: Overview, Insight and Visualisation Dataset

Investigation	Insight
Konten Corellation 3 Feature with the Target	The three features mentioned above, PaidStarPerHour, Broadcast Hours, and Gender, show a strong correlation with the potential for higher PaidStarPerHour based on the gender and the duration of their broadcasts.



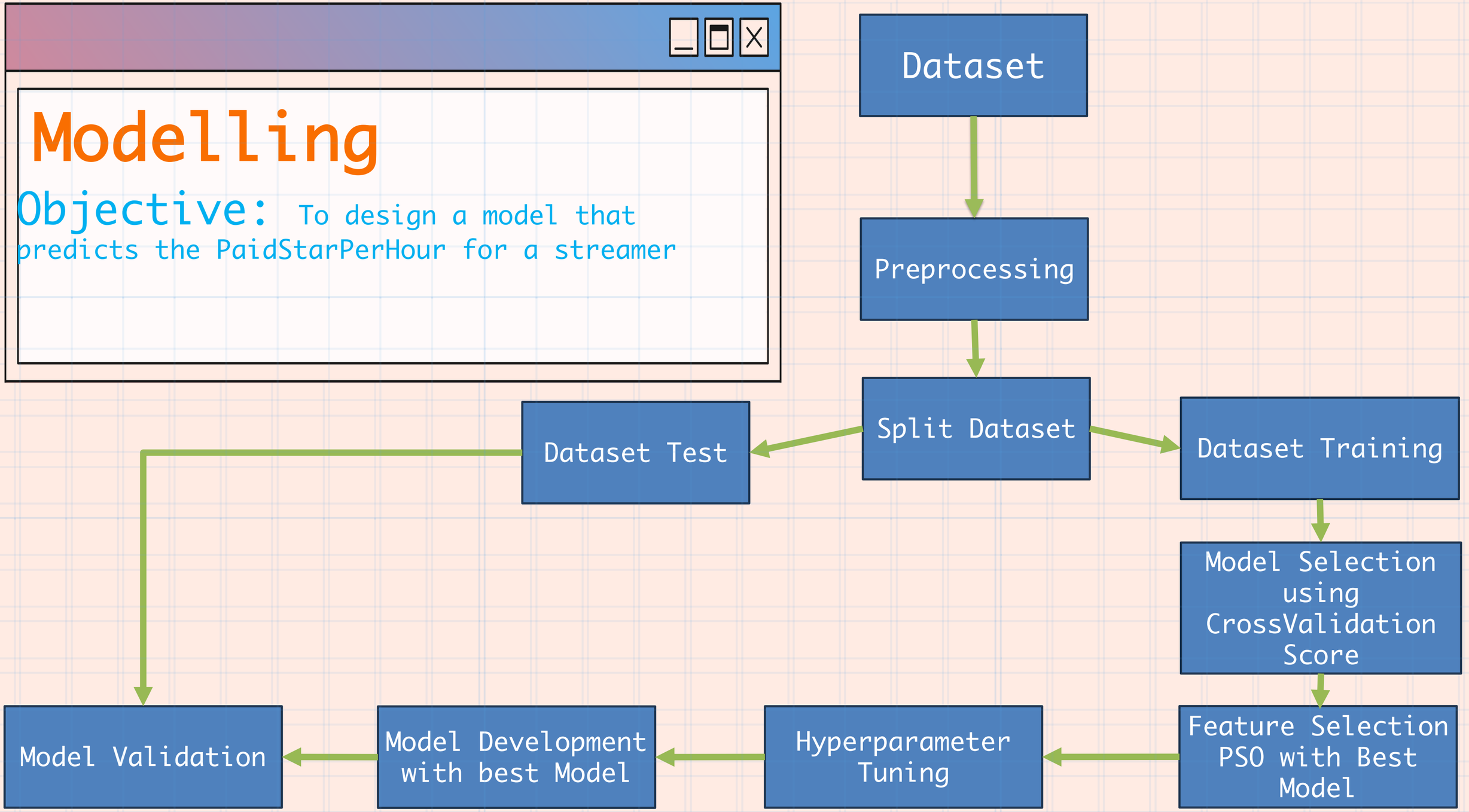
Preprocessing

Objective: Preparing data for Analysis and Modelling by removing inrelevant feature

Preprocessing	Function	Input	Output
Feature Selection	Drop inrelevant features from the dataset	Data_train	we have 29 highly relevant features in determining PaidStarPerHour."
Convert To Numeric	Label Encoding	String	Numeric
Handle Inrelevant Data	Replace Data Inrelevant to Relevant Data	Data Inrelevant Data	Relevant Data
Feature Scalling	StandarScaller	Data_train and Data_test	This technique ensures that the data is centered around 0 with a consistent standard deviation.

Modelling

Objective: To design a model that predicts the PaidStarPerHour for a streamer



Modelling Score

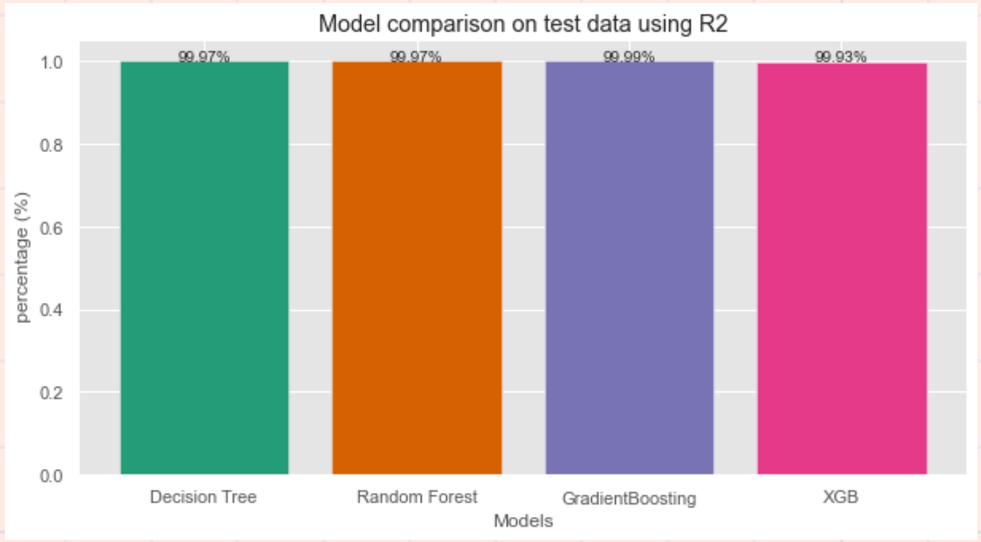
Objective: Design a model that predicts whether a household churns or is kept

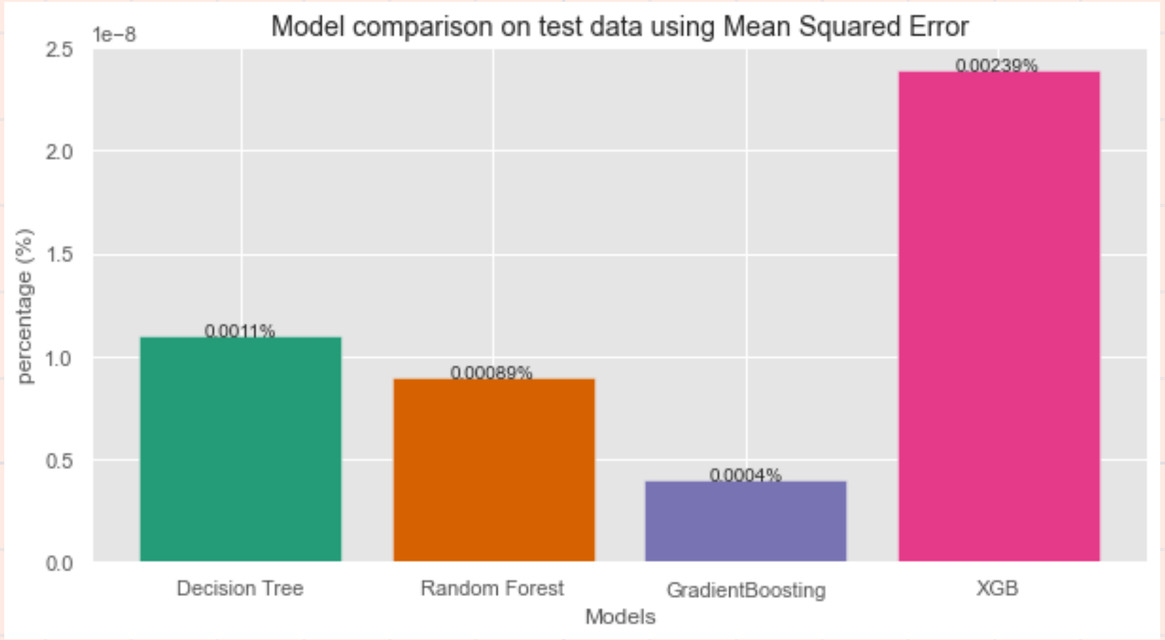
From the comparison of 7 models, it has been found that the best-performing model is the Gradient Boost Regressor (based on Cross Validation Score) with a score of 99%.

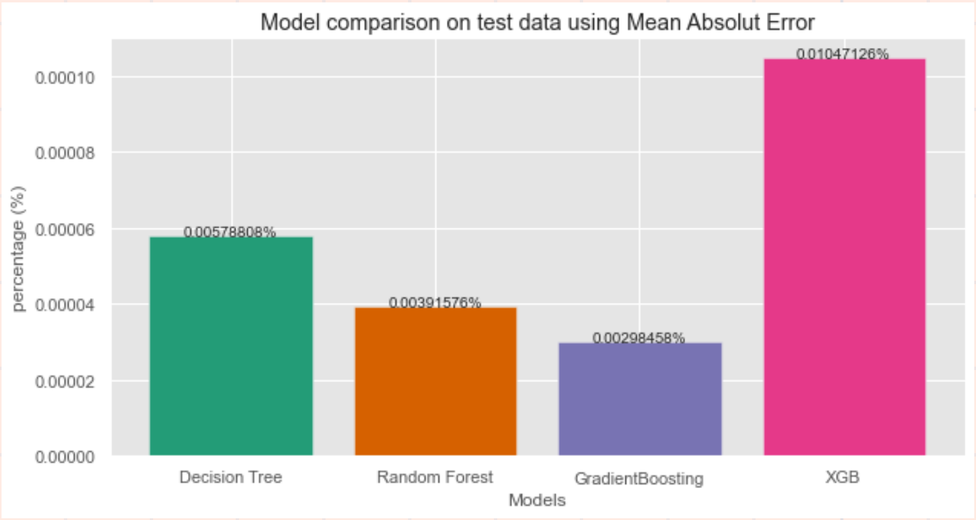
Algorithm	Cross Validation Score
Random Forest	0.9993525656112665
KNN	0.046330007414655136
Logistic Regression	0.9985964509549792
Decision Tree Regressor	0.9985964509549792
XGBoost Regressor	0.9948351621921558
Gradient Boost Regressor	0.9998108971673627
Support Vector Regressor (Kernel : RBF)	0.5384543285922325

Test set Model

Objective: Regression Report for each model



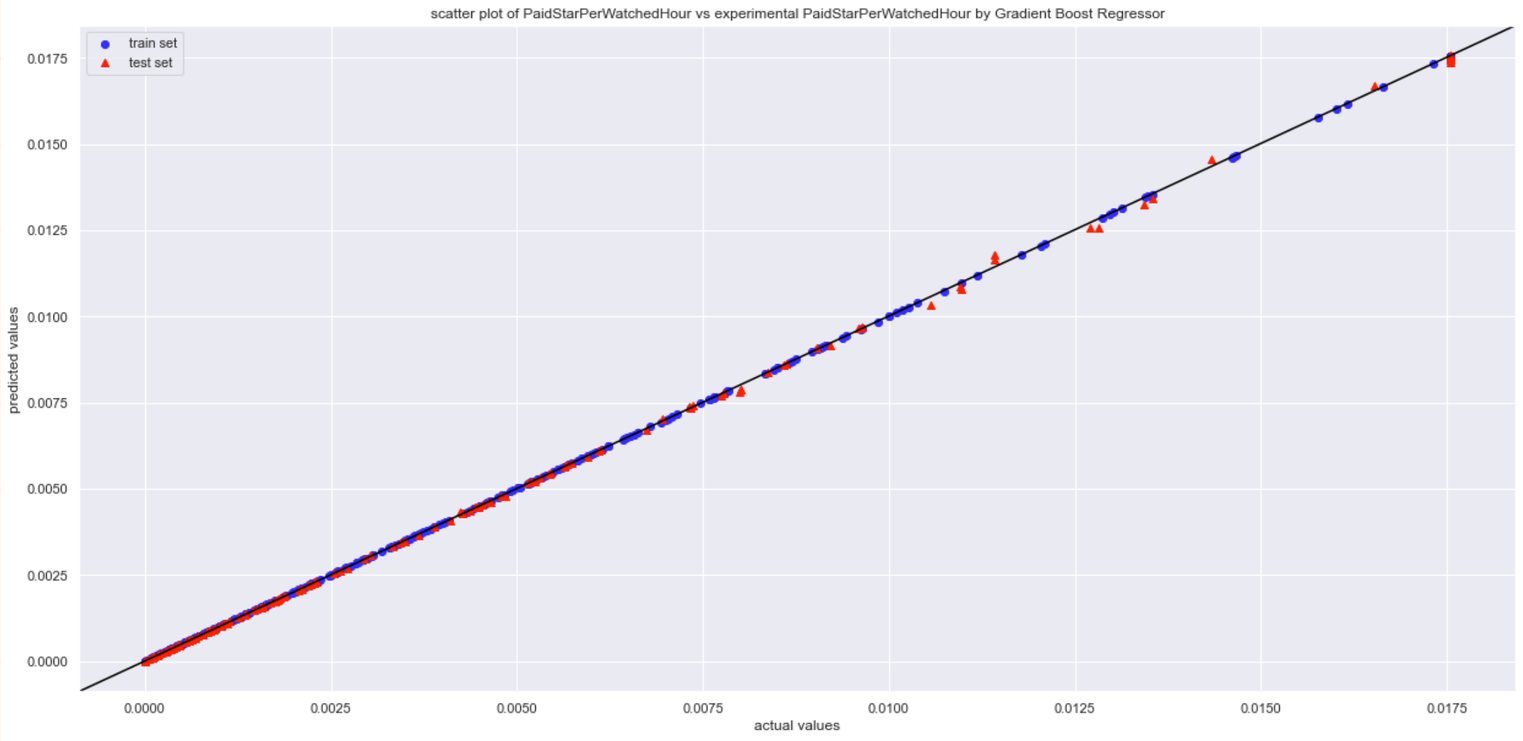




Berdasarkan hasil dari Grafik Regression report kami mendapatkan nilai yang sangat baik dimana nilai dari test set mendapatkan nilai metrik r2 dengan nilai 99% yang dimana merupakan nilai yang tinggi dalam mentukan hubungan antar tiap feature pada sebuah model, juga pada metrik MSE dan MAE mendapatkan nilai yang sangat kecil yaitu 0.0004% dan 0.0029% yang dimana kemungkinan terjadinya error sangat kecil pada model yang dibangun

Predict Model

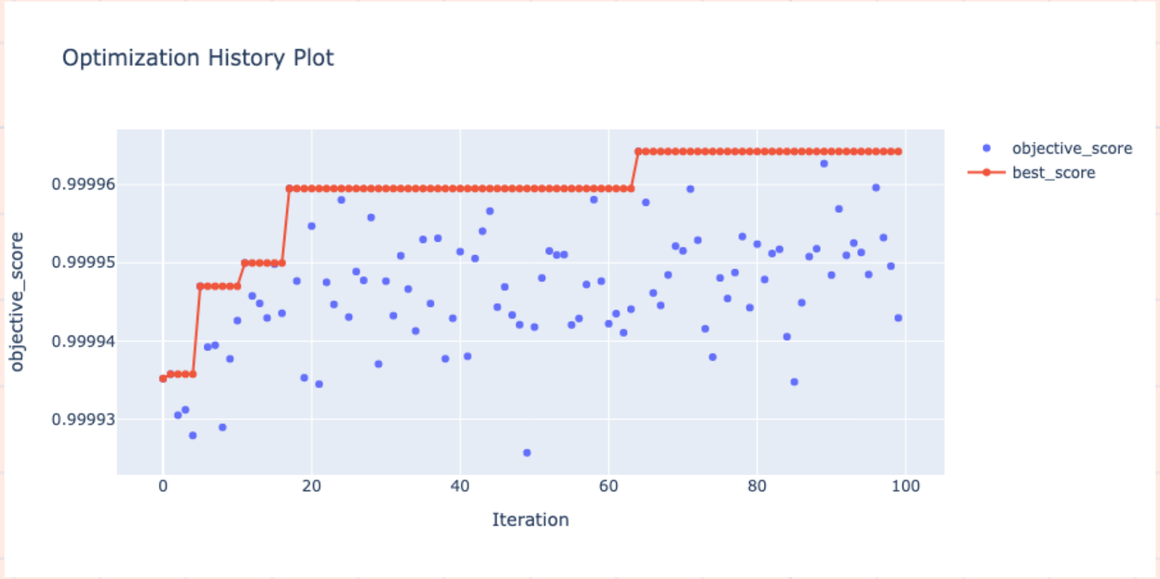
Objective: Implementation Model by unlabel data



Grafik Scatter Plot disamping menunjukkan bukti bahwa nilai prediksi dan nilai aktual sejalan serta memiliki nilai R2_Score 99% dan jumlah MSE dan MAE yang sangat rendah

Summary

In the context of determining the features that influence the target variable PaidStarPerHour for each streamer, we have performed feature scoring using the PSO-GBR algorithm. The algorithm randomly selected candidate features 100 times, and we obtained a score of 99%.



Based on the analysis with feature scoring using PSO-GBR, there are several factors that can affect PaidStarPerHour. Top 5 features that can have a positive impact on PaidStarPerWatchedHour are:

1. Personal_Values_Facet_Cont_Hedonism: this characteristic can indicate how much viewers enjoy the content they watch. Therefore, showing more enjoyable content for viewers can increase PaidStarPerWatchedHour
2. Temperament_Diligent: this characteristic can indicate how diligent and persistent content creators are in making high-quality videos. The more diligent and persistent content creators are in making high-quality videos, the higher PaidStarPerWatchedHour they can obtain.
3. Role_Director: content creators who act as directors tend to have greater control over the videos produced, so they can influence the quality of the videos produced and ultimately the PaidStarPerWatchedHour obtained.
4. Character_Facet_Cont_SelfConsciousness: this characteristic can indicate how confident content creators are in speaking in front of the camera and presenting themselves in front of the public. The more confident content creators are, the easier it is for them to attract the attention of viewers and ultimately increase PaidStarPerWatchedHour
5. Personal_Values_Facet_Cont_UniversalismConcern: this characteristic can indicate how much content creators consider the interests of society as a whole in the content they create. Content that attracts the attention of the general public can increase PaidStarPerWatchedHour

Suggestions

Here are some suggestions to address the issue of highly correlated features with PaidStarPerHour for streamers :

Based on the mentioned features, here are some suggestions to increase PaidStarPerHour:

1. Enhance Entertaining Content: Based on Personal_Values_Facet_Cont_Hedonism, it is important to focus on creating content that is enjoyable and entertaining for viewers. Provide content that offers engaging and positive experiences for the audience. This can increase viewer engagement and potentially boost PaidStarPerHour.

2. Improve Video Quality: In relation to Temperament_Diligent, encourage content creators to be more diligent and persistent in creating high-quality videos. Focus on production aspects such as lighting, sound quality, effective editing, and valuable content. High-quality videos will attract more viewers and potentially increase PaidStarPerHour.

3. Take Control as a Director: If content creators have the role of a director (Role_Director), encourage them to take greater control over the produced videos. By exercising creative control, editing, and storytelling, they can create more compelling and unique content. This can enhance viewer interest and increase PaidStarPerHour.

4. Boost On-Camera Confidence: Addressing Character_Facet_Cont_SelfConsciousness, provide training and support to increase content creators' confidence when appearing on camera. Communication, presentation, and public speaking training can help them feel more confident and capture the audience's attention effectively, leading to a potential increase in PaidStarPerHour.

5. Consider General Public Interests: In the context of Personal_Values_Facet_Cont_UniversalismConcern, it is essential to understand the interests and preferences of the general public. Encourage content creators to create content that is relevant and appealing to a broader audience. Understanding the interests of the general public can attract more viewers and potentially increase PaidStarPerHour.

By implementing these suggestions, the company can help content creators improve their PaidStarPerHour by leveraging features that have a positive impact.