

Trending Youtube Video Prediction in US

Sully Vo _ Team 1

How can the new Youtubers get more views from their video?

Few problems to increase the views :

tags, keys, contains...

Dataset : Kaggle



Stakeholders

- New Youtubers with Animal videos through 2017-2018 to predict the factors that helps them increase the views
- Youtube company



Step by Step Machine learning Algorithm

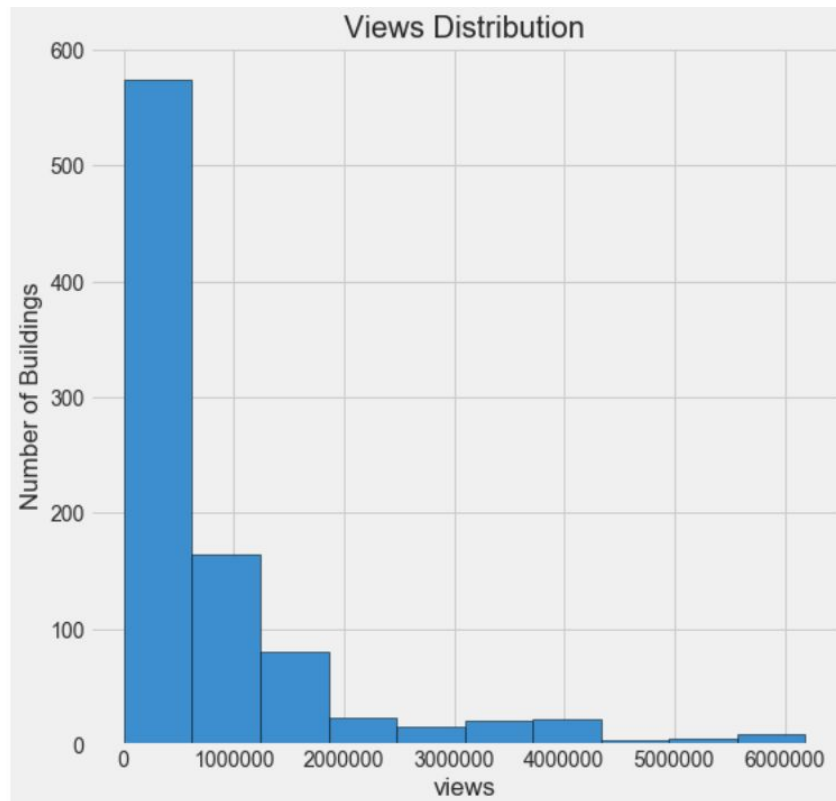
1. Data cleaning and formatting
2. Exploratory data analysis
3. Feature engineering and selection
4. Compare several machine learning models on a performance metric
5. Perform hyperparameter tuning on the best model

Data Exploration

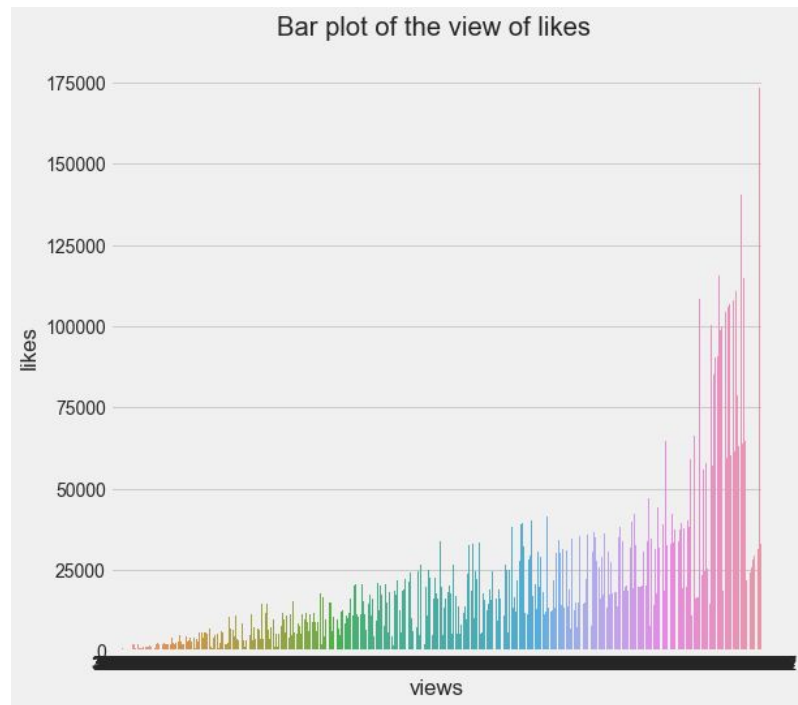
- Original data : 40949 rows
- 920 rows of data with 16 columns
- Interested in Animal channel (categorical_ID = 15)
-

	category_id	views	likes	dislikes	comment_count
count	920.0	920.0	920.0	920.0	920.0
mean	15.0	831143.0	21055.0	573.0	2892.0
std	0.0	1102091.0	25425.0	779.0	4842.0
min	15.0	3393.0	6.0	0.0	0.0
25%	15.0	185072.0	5433.0	108.0	459.0
50%	15.0	444502.0	14432.0	276.0	1173.0
75%	15.0	941299.0	26694.0	677.0	3270.0
max	15.0	6187457.0	178243.0	4899.0	44063.0

Data Exploration

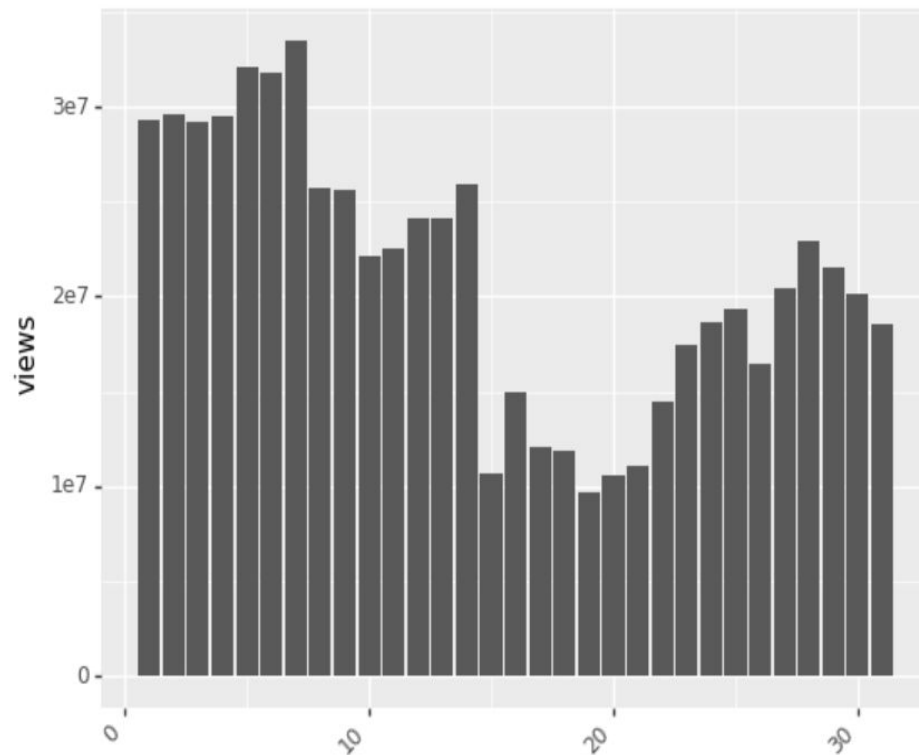


Views and likes

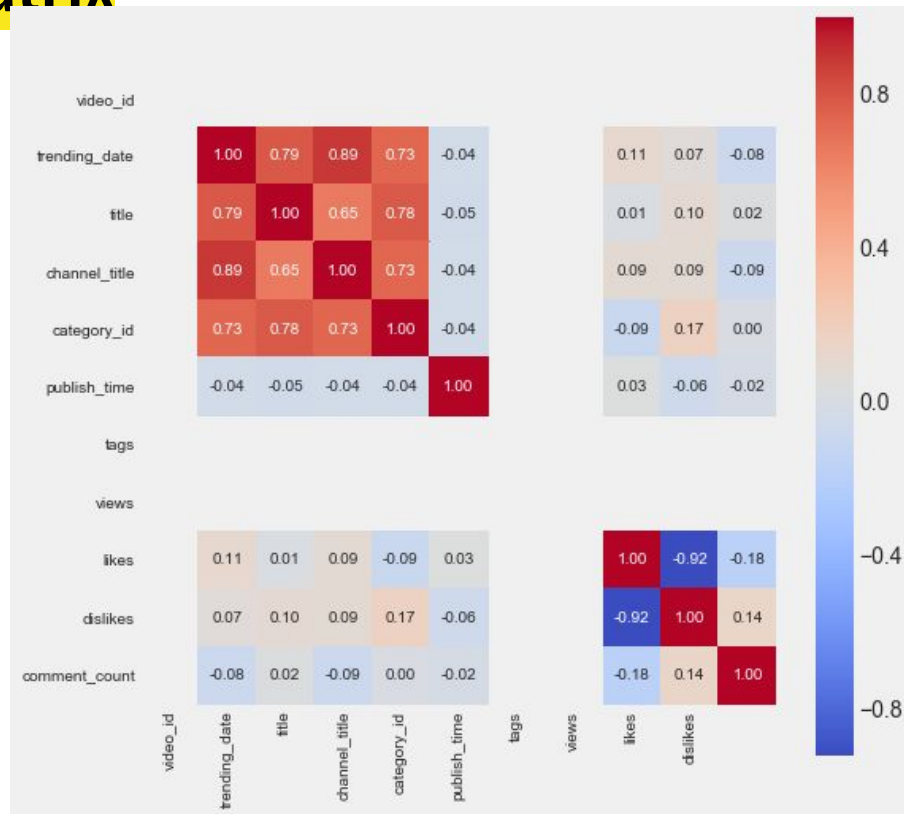


Months and Views

Plot the Months and Views of Animal Video in California in 2018 in US



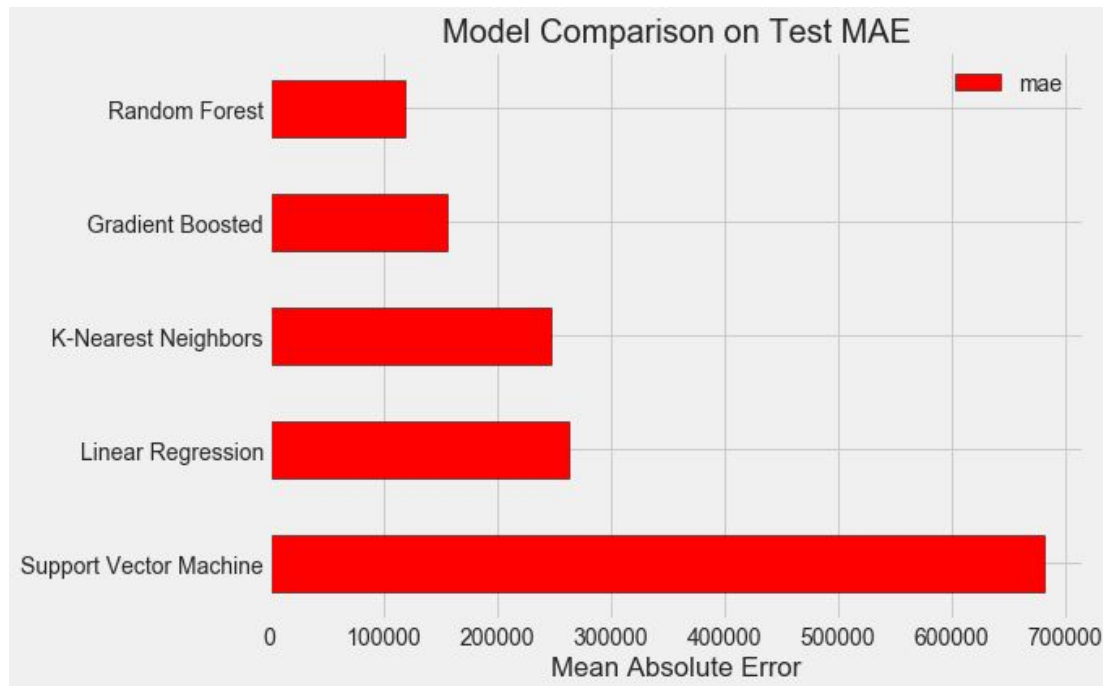
Correlation matrix



Feature engineering and selection

- Splitting data into training and testing
- Apply scaling
- Apply models :
 - Linear Regression
 - Support Vector Machine Regression
 - Random Forest Regression
 - Gradient Boosting Regression
 - K-Nearest Neighbors Regression
-

Comparing and choose the best model



Tuning part with Hyperparameter and Cross validation

Evaluate the default model

Model Performance

Average Error: 114351.7170 degrees.

Accuracy = 69.86%.

Evaluate the best random search model

Model Performance

Average Error: 37125.8181 degrees.

Accuracy = 91.28%.

Conclusion and Recommendation

- Collect more data as features
- Try different other model
- Apply model on testing set ,other data or only with importance features

Recommendation:

- YouTube's bot detection capabilities are getting better and better
- The algorithms are paying more attention to user behavior rather than view counts