Northeastern University

Decoding: What Drives Super Bowl Ad Success

Nishith Reddy, Zachary Armand, Zachary Becker, Tejas Mohan Karkera
DS 5220 Supervised Machine Learning Fall 2023

Introduction:

We are seeking to find out what the driving force is behind the popularity of Super Bowl commercials. The Super Bowl provides an opportunity for brands to promote their products with unique and often over-the-top commercials, so we want to discover what the defining characteristics are for a successful Super Bowl ad.

Brands pay millions of dollars for an ad during Super Bowl commercial breaks, and a significant portion of viewers enjoy the commercials just as much as the actual sporting event. Viewers often rank their favorite commercials and brands can truly make significant profits if they can put out a hit commercial. Thus, it is valuable to see which combination of absurdity, hilarity, danger, sex appeal, etc. enthralls the most viewers.

Our dataset consists of 247 advertisements from the 10 brands that aired the most commercials this century. Each row represents a different advertisement. Columns describe different variables relating to each advertisement. Variables range from binary descriptors (funny, patriotic, etc.) to video performance and descriptors (view count, like count, comment count, video title, channel title).

	year	view_count	like_count	dislike_count	favorite_count	comment_count	category_id
count	247.000000	2.310000e+02	225.000000	225.000000	231.0	222.00000	231.000000
mean	2010.190283	1.407556e+06	4146.031111	833.537778	0.0	188.63964	19.316017
std	5.860872	1.197111e+07	23920.402935	6948.521873	0.0	986.45691	8.004328
min	2000.000000	1.000000e+01	0.000000	0.000000	0.0	0.00000	1.000000
25%	2005.000000	6.431000e+03	19.000000	1.000000	0.0	1.00000	17.000000
50%	2010.000000	4.137900e+04	130.000000	7.000000	0.0	10.00000	23.000000
75%	2015.000000	1.700155e+05	527.000000	24.000000	0.0	50.75000	24.000000
max	2020.000000	1.763734e+08	275362.000000	92990.000000	0.0	9190.00000	29.000000

Fig 1. Statistical behavior of the data

Some of our columns have null values, as shown below. Thumbnail and description are in the lead with 129 and 50 null values, respectively. We decided to drop these columns since they are not relevant to our modeling.We then changed the boolean columns given below, funny, patriotic, danger, etc. into type integer to make it easier for our analysis.

```
year                        0
brand                       0
superbowl_ads_dot_com_url   0
youtube_url                11
funny                       0
show_product_quickly        0
patriotic                   0
celebrity                   0
danger                      0
animals                     0
use_sex                     0
id                         11
kind                       16
etag                       16
view_count                 16
like_count                 22
dislike_count              22
favorite_count             16
comment_count              25
published_at               16
title                      16
description                50
thumbnail                 129
channel_title              16
category_id                16
dtype: int64
```

```
year                           int64
brand                         object
superbowl_ads_dot_com_url     object
youtube_url                   object
funny                           bool
show_product_quickly            bool
patriotic                       bool
celebrity                       bool
danger                          bool
animals                         bool
use_sex                         bool
id                            object
kind                          object
etag                          object
view_count                   float64
like_count                   float64
dislike_count                float64
favorite_count               float64
comment_count                float64
published_at                  object
title                         object
channel_title                 object
category_id                  float64
dtype: object
```

Fig 2. Null Value analysis and data type given

Wrapping up our introduction to the data, we show a description of the columns to see some of the applicable statistics. This information gives us a very clear insight of what plots and the analysis which can be done so that we can understand the modeling analysis necessary for our data.

	year	funny	show_product_quickly	patriotic	celebrity	danger	animals	use_sex	view_count	like_count	dislike_count	favorite_count	comment_count	category_id
count	231.000000	231.000000	231.000000	231.000000	231.000000	231.000000	231.000000	231.000000	2.310000e+02	231.000000	231.000000	231.0	231.000000	231.000000
mean	2010.216450	0.688312	0.696970	0.173160	0.298701	0.307359	0.380952	0.259740	1.407556e+06	4038.406926	812.116883	0.0	181.329004	19.316017
std	5.766914	0.464189	0.460566	0.379207	0.458682	0.462402	0.486675	0.439444	1.197111e+07	23615.574169	6858.549918	0.0	967.648474	8.004328
min	2000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000e+01	1.000000	1.000000	0.0	0.000000	1.000000
25%	2005.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	6.431000e+03	14.500000	1.000000	0.0	1.000000	17.000000
50%	2010.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	4.137900e+04	120.000000	6.000000	0.0	9.000000	23.000000
75%	2015.000000	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.700155e+05	480.500000	24.000000	0.0	45.500000	24.000000
max	2020.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.763734e+08	275362.000000	92990.000000	0.0	9190.000000	29.000000

Fig 3. Final Statistical behavior of the data

Exploratory Data Analysis (EDA)

In our exploratory data analysis (EDA), our primary focus was directed towards examining binary features like funny, show_product_quickly, patriotic, etc., as these are the features that help us to answer the research question.

We plotted the distributions of boolean columns in our dataset for our first visualization. The x-axis is labeled with the names of the boolean features, while the y-axis represents the count. Within this plot, a value of 0 signifies the absence of a specific feature in the advertisement, while a value of 1 indicates the presence of that particular feature. With the help of this visualization, the most commonly occurring and least prevalent features can be identified.

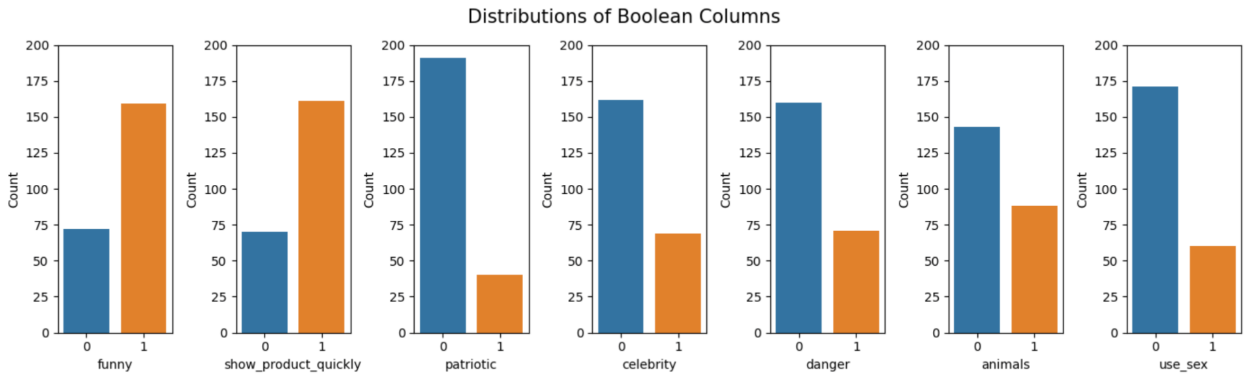

Fig 4. Distribution of Boolean Columns

For the second plot, we visualized a stacked bar plot with the count of advertisements over the years and subsequently visualized the proportion of binary features in the dataset for each year. The x-axis is labeled with the years from 2000 to 2020, and the y-axis gives the count of advertisements. The thickness of the distinctive hues within each bar signifies the proportion of binary features embedded in the dataset for the corresponding year. This graph facilitates the identification of changes in the trends of the advertisements over the years.

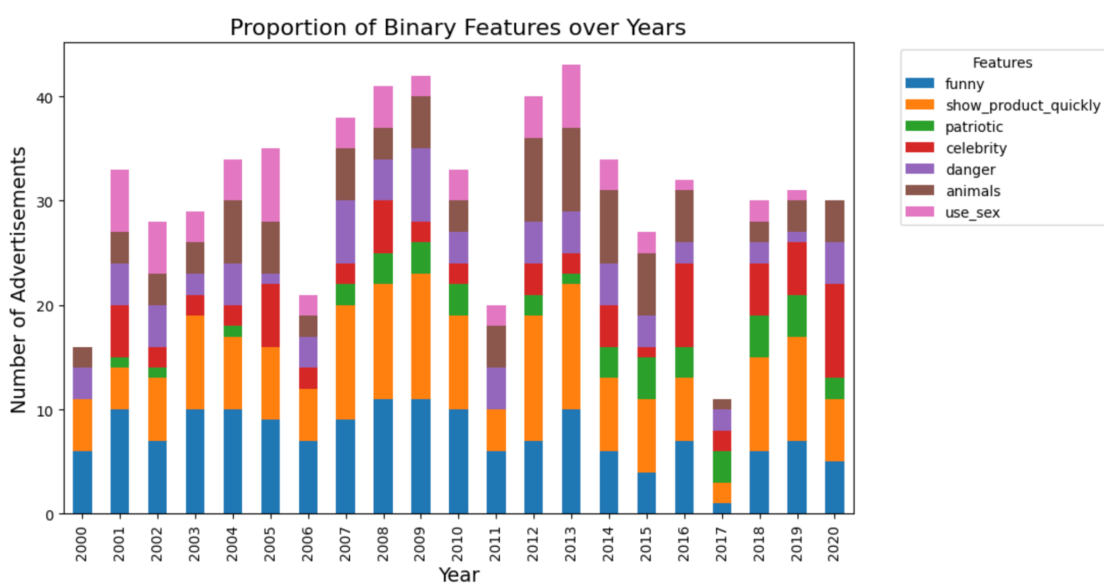

Fig 5. Proportion of Binary Features Over the Years

Lastly, we categorized the brands broadcasting their advertisements in the Super Bowl into their product categories and counted the total number of advertisements per category. After getting the count for each category, a pie chart is plotted. Although this pie chart is not very handy in answering our research question, it gives an idea about what kind of brands are mostly interested in broadcasting their advertisements in the Super Bowl. From the picture below, most advertisements in the Super Bowl are from beverages/snacks companies.

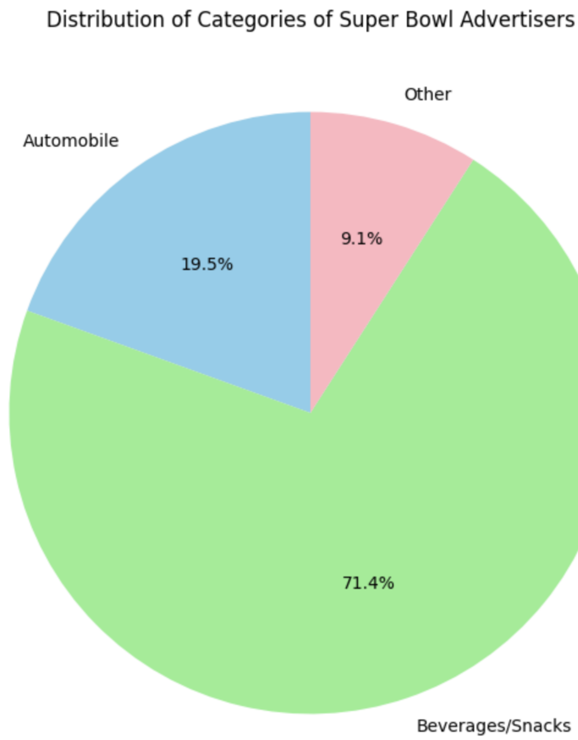

Fig 6. Distribution of Categories of Super Bowl Advertisers

Target and Features

Our features include seven binary columns describing whether a video included a particular characteristic. For example, if a video included humorous elements, it would have a value of 1 in the field "funny". We included these seven columns as predictors in order to see which video elements influence a video's popularity. The full list of columns include whether a video included: funny elements, showed a product quickly, patriotic elements, celebrity, dangers, animals, or used sex appeal. By using these inherent qualities of a video, we can hopefully determine which combination of elements leads to a higher measure of popularity. We dropped columns that were irrelevant to our research question or were unable to provide valuable insights using the models that we chose (such as video title, YouTube channel name, etc.). None of these binary columns included NA values.

Our target feature therefore aimed to quantify popularity based on some features of the videos in the dataset. We decided to create a metric of popularity by multiplying the number of views a video had by the like-to-dislike ratio of that video, making sure to replace any NA values with 1 before performing our calculation to avoid dividing by zero.

$$Popularity\ (p)\ =\ Views\ (v)\ *\ (Likes\ (l)\ /\ Dislikes\ (d))$$

By multiplying the number of views by the like-to-dislike ratio, we can further quantify popularity by rewarding highly liked videos and penalizing heavily disliked videos, which might indicate a poorly received yet highly viewed commercial. We then determined the 80th percentile based on this metric, and categorized a video as "Popular" if it was above the 80th percentile, and "Not Popular" if it was below the 80th percentile. The distribution of the "Popularity" metric and the 80th percentile can be seen in the figure below:

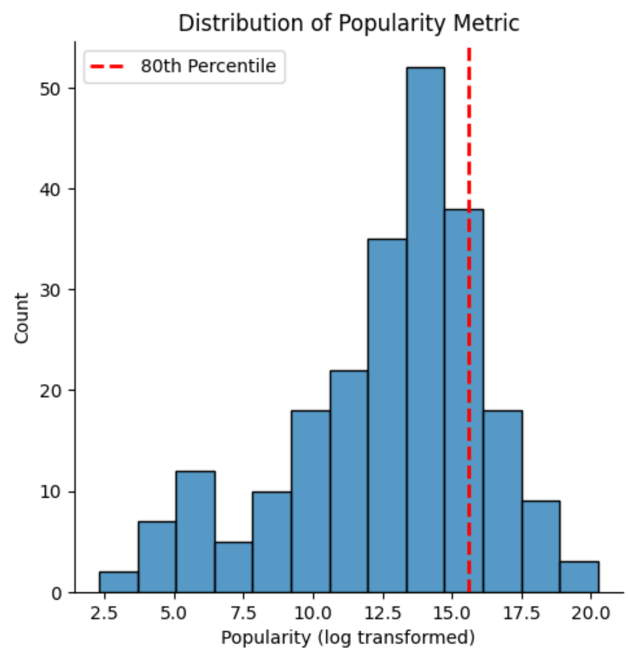

Fig 7. Distribution of Popularity Metric

This threshold for popularity captures the highest-performing videos and captures any outliers, but still balances the ratio of videos considered popular.

Models

Modeling a classification task is heavily dependent on the kind of feature selection and quality of features we want to select. For this specific task, we have identified the most informative features using the Exploratory data analysis and plots previously drawn. For the purpose of modeling and model selection we have identified Logistic regression and Random Forest as the two most important models for our task. Mainly the reasons are as follows:

Why Logistic Regression?

- Given our binary classification, logistic regression is a great way to determine popular or not.
- Also allows for simple determination of feature importance through coefficients.
- Logistic Regression is easy to interpret.

Why Random Forest?

- Random forests will give high predictive accuracy and can handle non-linear relationships in our data.
- We will get a measure of feature importance which allows for measuring which features contribute the most to the model's predictions.
- Also fine-tuned random forest to choose best hyperparameters.

Model Results

After careful consideration towards the performance of the models and analyzing the plots I feel that there is a lot to with the choice of the thresholds and parametric usage in the models and for this purpose we had used the box plot of the log transformed values and with this we had performed the fine tuning for the random forest model. The results for the model are as follows:

Logistic Regression was able to generalize to some extent and with performance of the model being 45%. Logistic regression was able to provide some features as the most important ones on the basis of the performance. With some of the top features being the ad being Funny, show_product_quickly, patriotic.

Classification Report (Logistic Regression):

	precision	recall	f1-score	support
Not Popular	0.80	0.42	0.55	38
Popular	0.19	0.56	0.28	9
accuracy			0.45	47
macro avg	0.49	0.49	0.41	47
weighted avg	0.68	0.45	0.50	47

Feature Importance:
funny: -0.399
show_product_quickly: 0.874
patriotic: 0.29
celebrity: 0.251
danger: 0.277
animals: -0.0737
use_sex: -0.703

Fig 8. Classification Report and Feature Importance from Logistic Regression

Random Forest Model was able to generalize much better then a logistic classifier and with performance of the model being 51%. Random Forest Model was able to provide some features as the most important ones on the basis of the performance. With some of the top features being the ad being show_product_quickly, animals, danger.

Classification Report (Random Forest):

	precision	recall	f1-score	support
Not Popular	0.80	0.53	0.63	38
Popular	0.18	0.44	0.26	9
accuracy			0.51	47
macro avg	0.49	0.49	0.45	47
weighted avg	0.68	0.51	0.56	47

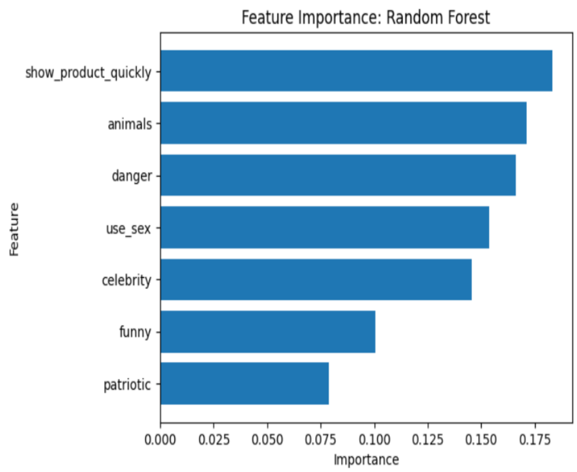

Fig 9. Classification Report and Feature Importance Plot from Random Forest

Random Forest Model Modified being our third model was able to generalize to a normal extent then a logistic classifier and previous random forest and with performance of the model being 64%. With some of the top features being the ad being patriotic, animals, funny. Random Forest Fine-Tuned or modified was able to provide the best possible performance and the parameters used were max_depth being None, max_features being sqrt and min_sample_leaf being 3 and n_estimators being 100.

Results

With emphasis on performance and model parameters we have observed some interesting insights from the model training and evaluation phase. What we have observed are some of the key insights which we would want to present below:

- Firstly because the thresholding approach is used it is providing some grounds to the fact that popularity when decided on the log transformed distribution shows that percentile can be a good contributor on deciding whether the ad would be popular or not popular.
- Our usage of 80 as the value of the threshold for popularity is providing quality results.
- The generalisability ability of the model can be still better and as tree based models are able to perform much better on imbalance data which is in our case we can see that random forest did a really good job specifically when it is fine tuned.
- The performance plot for the models can be seen below wherein the performance of the final random forest model fine tuned was the best.

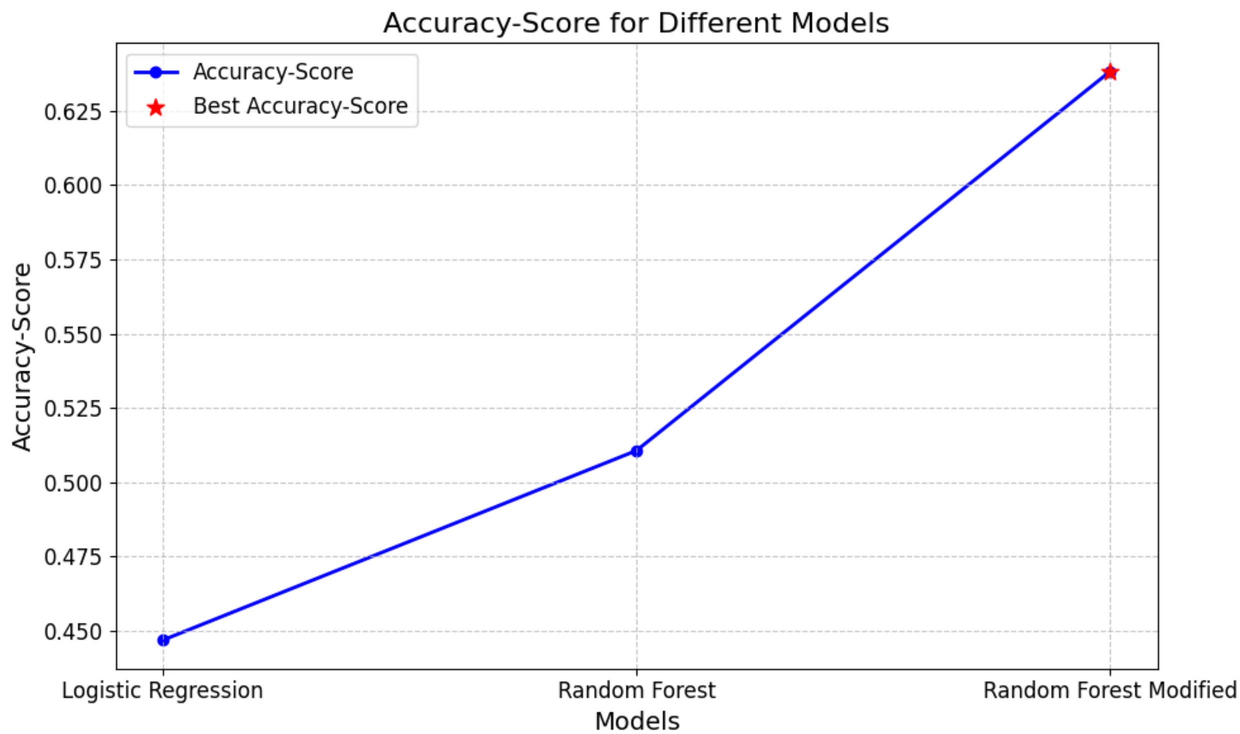

Fig 10. Accuracy Score for Different Models

Conclusion

Target for our research problem is defined by “View count * (Like count / Dislike count)” and assigning popular tags using a thresholding approach is something which can be considered when we are trying to consider an ad being popular or not being popular. We can see that popularity is a measure of some core features when it comes to the Super Bowl ads like the patriotic, animals, and funny. This showcases that these are something we need to always consider when deciding the popularity of the Super Bowl ads.