Predicting Eurovision Finalists

User Guide

Copyright Information

Table of Contents

Preface and General Information

This is my capstone project for my Data Science MS degree at Grand Canyon University. The goal of this project is to build models to predict which Eurovision semi-finalists will qualify for the finals. Eight different models have been built using data from the years 2012-2021 with the variables Wiwi Jury Score, OGAE Score, YouTube Views, YouTube Like Percentage, Spotify Listens, English vs Non-English Language, Pop vs Not Pop, Dance, Energy, Live, Acoustic, Happy, Speech, Loud, and Tempo.

The models that were built are linear regression (OLS), reduced variable linear regression (OLS-Red), logistic regression (Log), reduced variable logistic regression (Log-Red), decision tree (Tree), random forest (RF), naive Bayes (NB), and ensemble. Models that output a probability rather than a binary value use .5 as the cut-off point between classification as qualifying versus non-qualifying.

System Summary and Getting Started

In order to view this dashboard, you must be using a desktop computer that has an internet connection and a web browser. If you intend to download the dashboard to your local machine or manipulate the Python code, your computer must be able to run the desktop version of Tableau Public and Python. For system requirements for these programs, please visit the following:

https://www.tableau.com/products/techspecs

https://www.python.org/downloads/operating-systems/

To use the Tableau dashboard or Python code, visit the following pages:

https://public.tableau.com/app/profile/kc.grace/viz/Baker-DataScienceCapstone/LandingPage?publish=yes

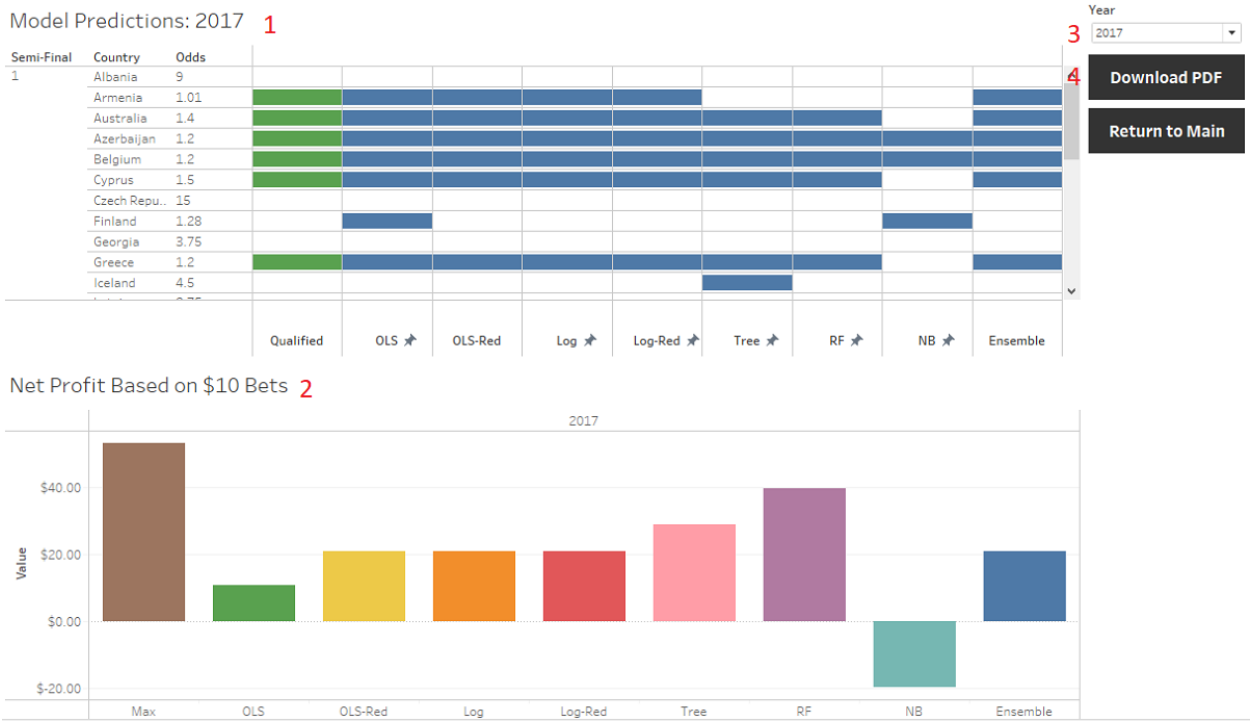
https://github.com/kcgb20/GCU-Capstone

Using the Tableau Dashboard

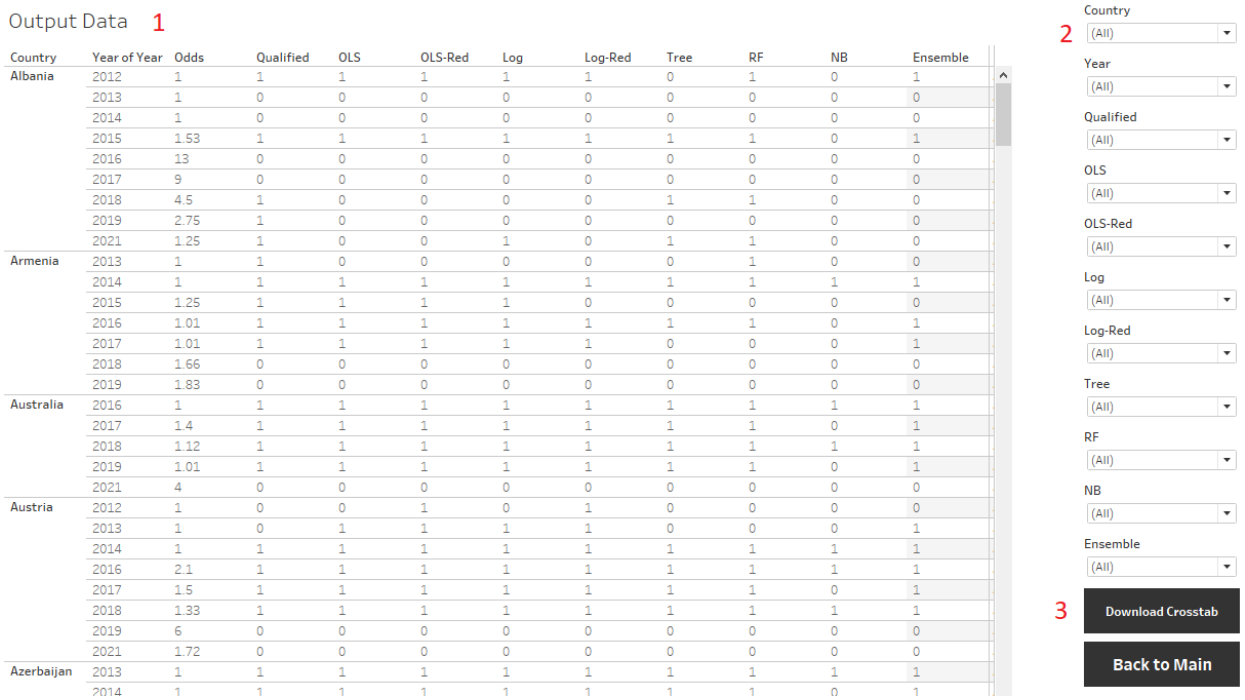
The Tableau Dashboard for this project gives information about the performance of the models. It is available on Tableau Public and can be downloaded for anyone to manipulate as they see fit. Below are the views that are built in to the pre-defined view.

The **Model Performance** page will give information about each individual models' performance. You can view these statistics by selecting a given year from the dropdown menu or you may view it in aggregate. A confusion matrix (1) is included to show true positives, false positives, false negatives, and true negatives. From these values we can derive the other metrics: accuracy, precision, and recall (2). You can read more about these metrics and how they are calculated here. You can filter the data by a given year (3) and download the visualization as a PDF (4).

The **Model Predictions and Betting Profits** page will show additional information about the models' predictions. The top graph (1) shows the contestant qualification as a green bar and predicted qualification as a blue bar by model; if a bar is not present, that means it was not predicted to qualify. You will also notice that the betting odds are shown on this graph as well; betting odds are available from years 2015 onward and were collected from Eurovision World. The bottom graph (2) uses these odds in conjunction with model prediction to determine net profit if a $10 bet was placed on each song the model predicted to qualify. The calculation for a true positive is [$10*(Odds-1)], a false positive is [-$10], and [$0] for true negatives and false negatives. The value "Max" has the theoretical returns if you placed bets on all qualifying songs with no incorrect bets. This view can be filtered by year (3) and downloaded as a PDF (4).



The **Model Output** (1) tab contains the information that feeds into this dashboard in a tabular format if you prefer text tables over graphs. You can filter this data by several different variables (2) and can download in a crosstab format (3).



I have also included a **Model Input** (1) tab so that you can see which variables went into the predictive models. This can be filtered by country or year (2) and can be downloaded in crosstab format (3).

Input Data 1

Year	Country	Song	Artist	Semi	Order	Odds	Qualified	Place	Returning	Youtube	Upvotes	Spotif
2012	Albania	Suus	Rona Nishliu	1	5	1	1	2	0	5	10	12
	Austria	Woki Mit Deim P..	Trackshittaz	1	16	1	0	18	0	8	16	6
	Belgium	Would You	Iris	1	8	1	0	16	0	16	17	13
	Cyprus	La La Love	Ivi Adamou	1	12	1	1	7	0	2	3	15
	Denmark	Should've Know..	Soluna Sam..	1	13	1	1	9	0	13	8	15
	Finland	Nar Jag Blundar	Pernilla Kar..	1	9	1	0	12	0	11	7	15
	Greece	Aphrodisiac	Eleftheria E..	1	3	1	1	4	0	4	12	4
	Hungary	Sound of Our He..	Compact Di..	1	15	1	1	10	0	9	2	7
	Iceland	Never Forget	Greta Salo..	1	2	1	1	8	1	3	1	5
	Ireland	Waterline	Jedward	1	18	1	1	6	1	10	15	2
	Israel	Time	Izabo	1	10	1	0	13	0	12	13	10
	Latvia	Beautiful Song	Anmary	1	4	1	0	16	0	17	11	14
	Moldova	Lautar	Pasha Parfe..	1	17	1	1	5	0	14	4	9
	Montenegro	Euro Neuro	Rambo Ama..	1	1	1	0	15	0	6	14	15
	Romania	Zaleilah	Mandinga	1	6	1	1	3	0	7	6	1

2

Country
(All)

Year
(All)

3

Download Crosstab

Return to Main

Using the Python Code

The Python code for the models can be found on GitHub along with the original data set for import. You can feel free to download the code and data to run the model as an exercise or even add new predictor variables to the underlying data set to see if you can create a better model. The modeling code is under Capstone Code.ipynb This code will create the following types of models:

-- Linear Regression

-- Linear Regression, reduced variable model

-- Logistic Regression

-- Logistic Regression reduced variable model

-- Decision Tree

-- Random Forest

-- Naive Bayes

-- Ensemble

The reduced variable models are based on including only variables that are significant from the full-variable models. The decision tree model is limited to 5 levels. The random forest model uses 1000 trees.

If you run this model you will need to change the path to where you put the downloaded data. If you add additional variables to the underlying data or remove them, you will also need to add or remove them to the definitions of *x, x_test, x_dec, x_dec_test, x_nb, x_nb_test, x_year, x_2_year, x_3_year, x_dec_year,* and *x_nb_year.* You will also want to review your significant variables from the linear regression and logistic regression models to input the correct variable selection

for *x_2, x_2_test, x_3,* and *x_3*_test. If you add years to the data, make sure to add running the predictions at the end of the code using the *runmodels()* function. If exporting the data, you will also have to add the year to *pd.concat* function.

Troubleshooting

This project has undergone extensive user testing which shows that there are no issues running the Python code or Tableau dashboard. The code runs without errors and Tableau displays information appropriately. If you are having errors when running the Python code, ensure that your file paths are pointing to the correct location of your data and that you have all packages installed to run the appropriate modules. You can easily search for documentation for any errors that Python may throw. If troubles occur with Tableau, you can consult Tableau's Help page:

https://www.tableau.com/support/help

For any other issues, please email me. Contact information is located at the end of this document.

FAQ

Here are some of the most frequently-asked questions.

Q: What makes your predictive model novel?

A: This model is one of the only models which predicts finalists. Most Eurovision predictive models only focus on predicting the contest winner. This model also takes into consideration various social media aspects.

Q: Why does your model only include data as far back as 2012?

A: Eurovision rules have changed drastically over time. It would not be appropriate to apply the model to years where rule sets are different.

Q: Why does your model only show betting gains for 2015 and later?

A: Reliable information for odds prior to 2015 was not available.

Q: Why is 2020 not included in your data?

A: Eurovision was canceled in 2020 due to COVID-19.

Q: Why doesn’t the dashboard work on my cellphone?

A: This dashboard was not optimized for cellphone view. If your cellphone has the ability to “View as Desktop”, you can use this to bypass the unoptimized cellphone view.

Contact Information and Help

For any comments, questions, or concerns, please email **kcgrace89@gmail.com**.

Legal Disclaimer

Predictions shown in this dashboard and produced by the models are not guarantees. The user accepts all risks when using these predictions. Creator is not to be held liable for any monetary gain or loss based on the use of predictions in this dashboard or any that come from the use and manipulation of the Python code.