Eurovision

SONG CONTEST

Introduction

The Eurovision Song Contest is an international songwriting competition organised annually by the European Broadcasting Union. Each participating country submits an original song to be performed on live television and radio to compete in a competition. Each country casts votes for others countries' songs to determine a winner.

Data

Our group recieved most of our data from the dataset www.github.com/-Spijkervet/eurovision-dataset. In addition to this, the Essentia music extractor was also used to analyse the audio of the songs to generate even more features from which to base our findings. For the lyric analysis, we used the NLP method to research frequent words, pairwise correlations and clusters of words. Features not deemeded necessary for analysis were columns like points and semi final data.

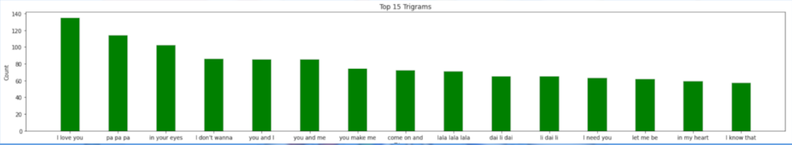
Models

Models used in our research include:
- Linear Regression
- Lasso Regression
- Ridge Regression
- Random Forest Classifier
- K Nearest Neighbour

From these, we measured the accuracy, and then used the most accurate one on the results of the 2020 contest

Objectives:

- To develop a method for predicting the results of ESC that is more accurate than just random guessing
- To analyse the lyrics of the songs in the contest and find patterns of words



Results

We found that the model with the highest range accuracy was the Random Forest Classifier, thus we decided to use this when predicting the results of the 2020 Contest. To the right you can view our predicted top 10 and bottom 10.

A sample of our lyric analysis results are seen above (Top 15 trigrams) and below (Wordcloud of most frequent words used)

Bulgaria	1
Slovenia	1
Switzerland	2
Spain	2
Georgia	3
Croatia	3
Belgium	5
Armenia	7
Belarus	7
North Macedonia	7

Germany	15
Serbia	15
Malta	16
Austria	16
Greece	16
Poland	18
Czech Republic	22
Italy	22
Latvia	22
San Marino	26

Conclusion

Due to the fact that our most accurate models only have an accuracy of at most 30%. I believe it is safe to assume that the results of ESC cannot be accurately predicted using the features we used. However, I do believe our model still produces results more accurate than just random guessing, considering that we still take into account many different features. We also recieved satisfactory results from our lyric analysis. I believe we achieved both of our objectives

Teele Tars Sander Soodla Maria Küüsvek