MASHABLE ARTICLE POPULARITY

What is Popularity?

- In this project, popularity was defined by the number of **shares** that an article received.
- As opposed to **views**, which represent the number of times an article was simply seen by a reader, **shares** imply that a reader viewed, read, and enjoyed an article so much they took the time to share it on social media for others to read.

What Makes Popularity?

In Short…

- Positive sentiment
- Long articles with simple words
- References
- Keywords, keywords, keywords!

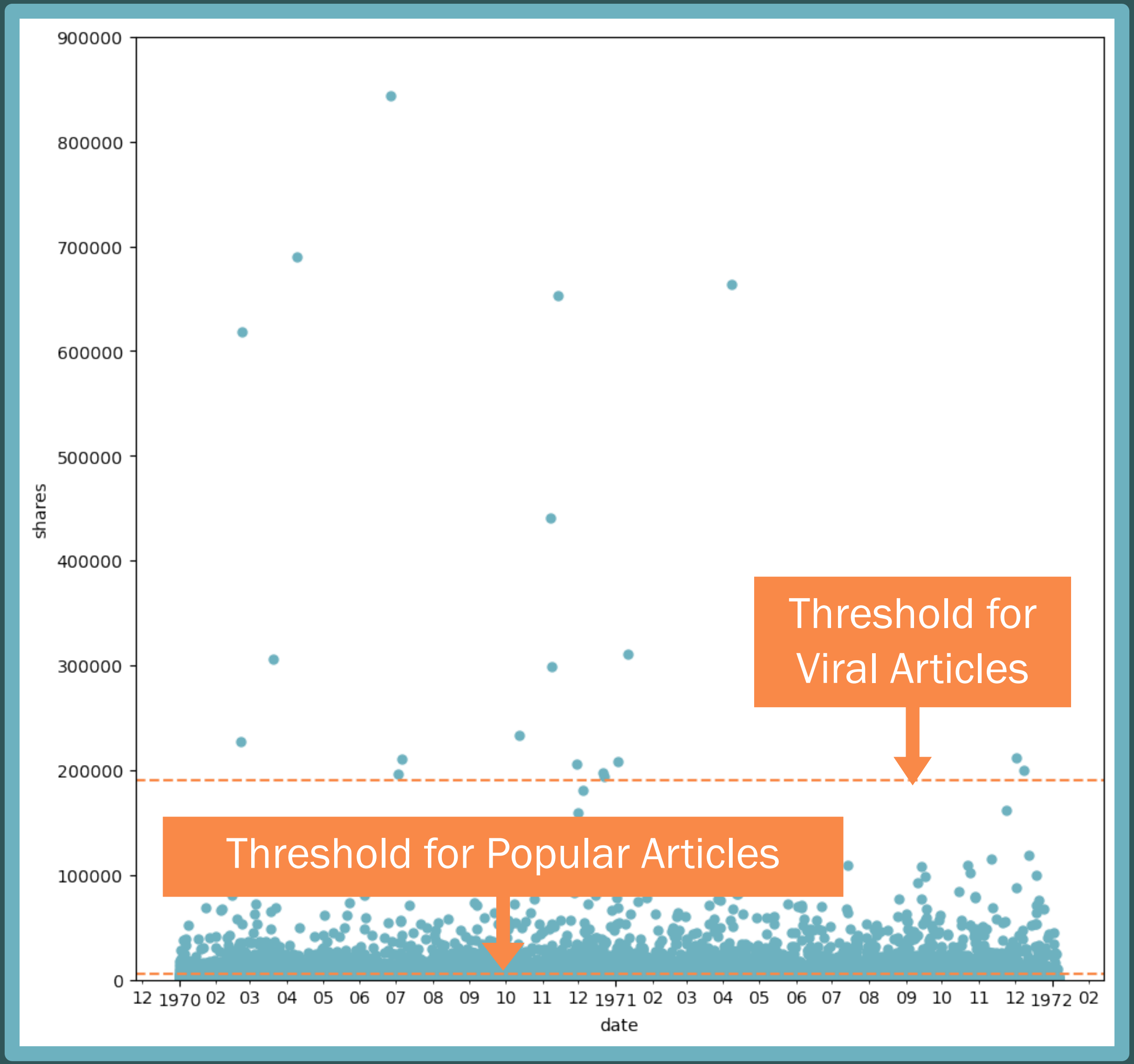
ANALYSIS

The Search for Popularity

Viral Articles

Viral articles are a subset of popular articles that preformed **significantly better** than other articles.

While most popular articles were under the 100,000 shares range, these viral articles hit as high as **800,000 shares.**

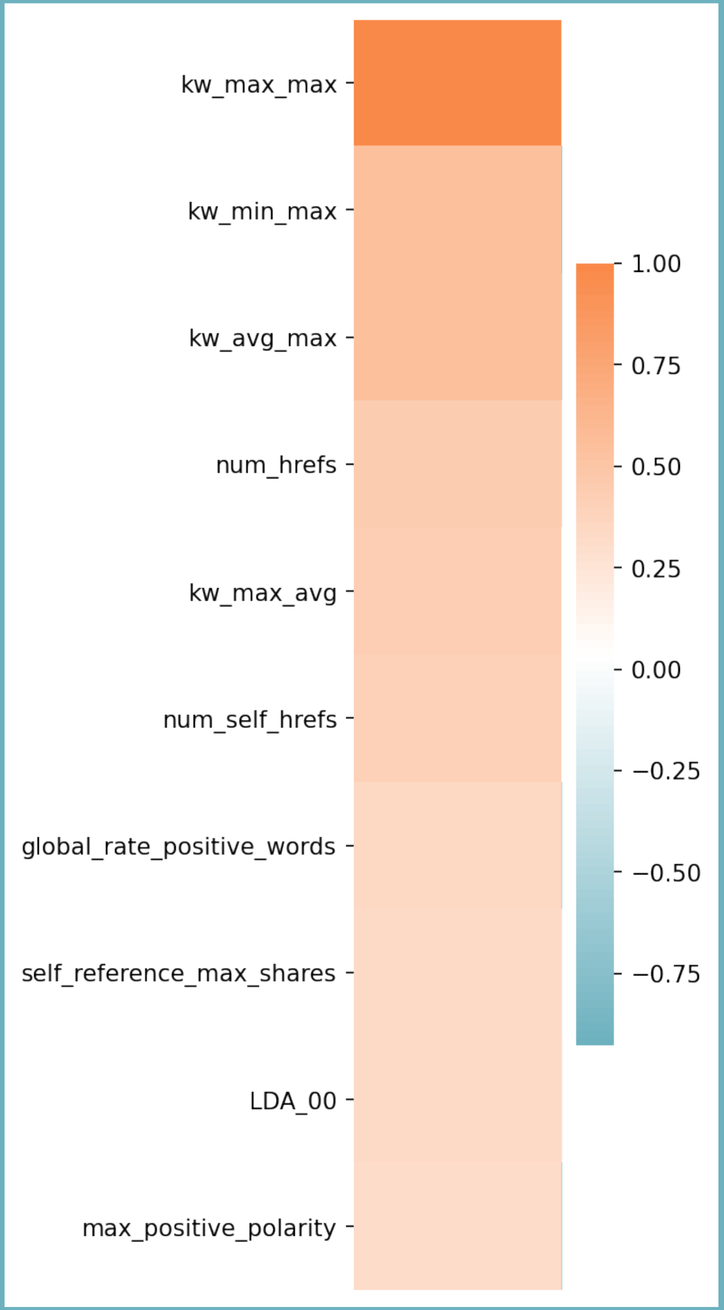


What Do Viral Articles Have in Common?

Separating the viral articles, I used a correlation matrix to determine what features correlated with shares.

Positive Correlation:

- High shares in all keywords
- References
- Positive Sentiment

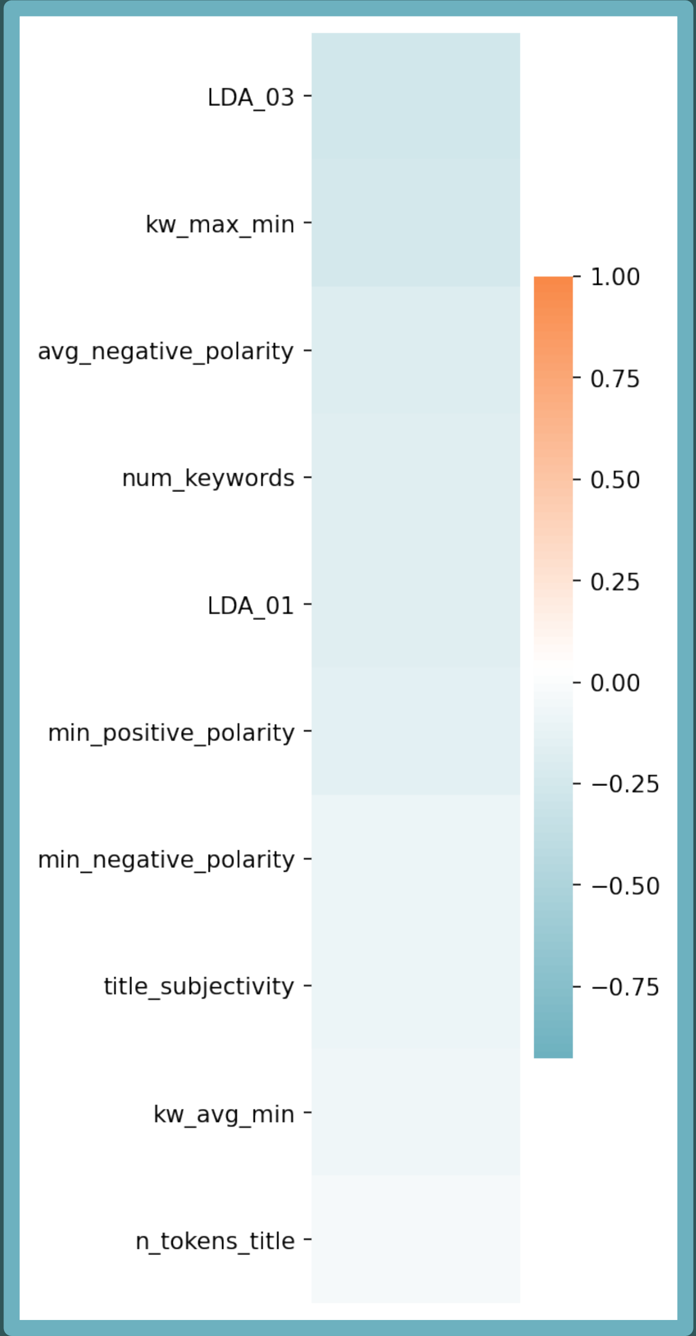


What Do Viral Articles Have in Common?

Separating the viral articles, I used a correlation matrix to determine what features correlated with shares.

Negative Correlation:

- Low shares in all keywords
- Negative Sentiment
- Title Subjectivity



Clustering

Using K-Means Clustering, the data was bunched into **four** separate groups.

Cluster 1

- Positive
- Less text, longer words

Cluster 2

- Positive and slightly subjective
- More text, shorter words
- More references

Cluster 3

- Negative
- Less text, longer words

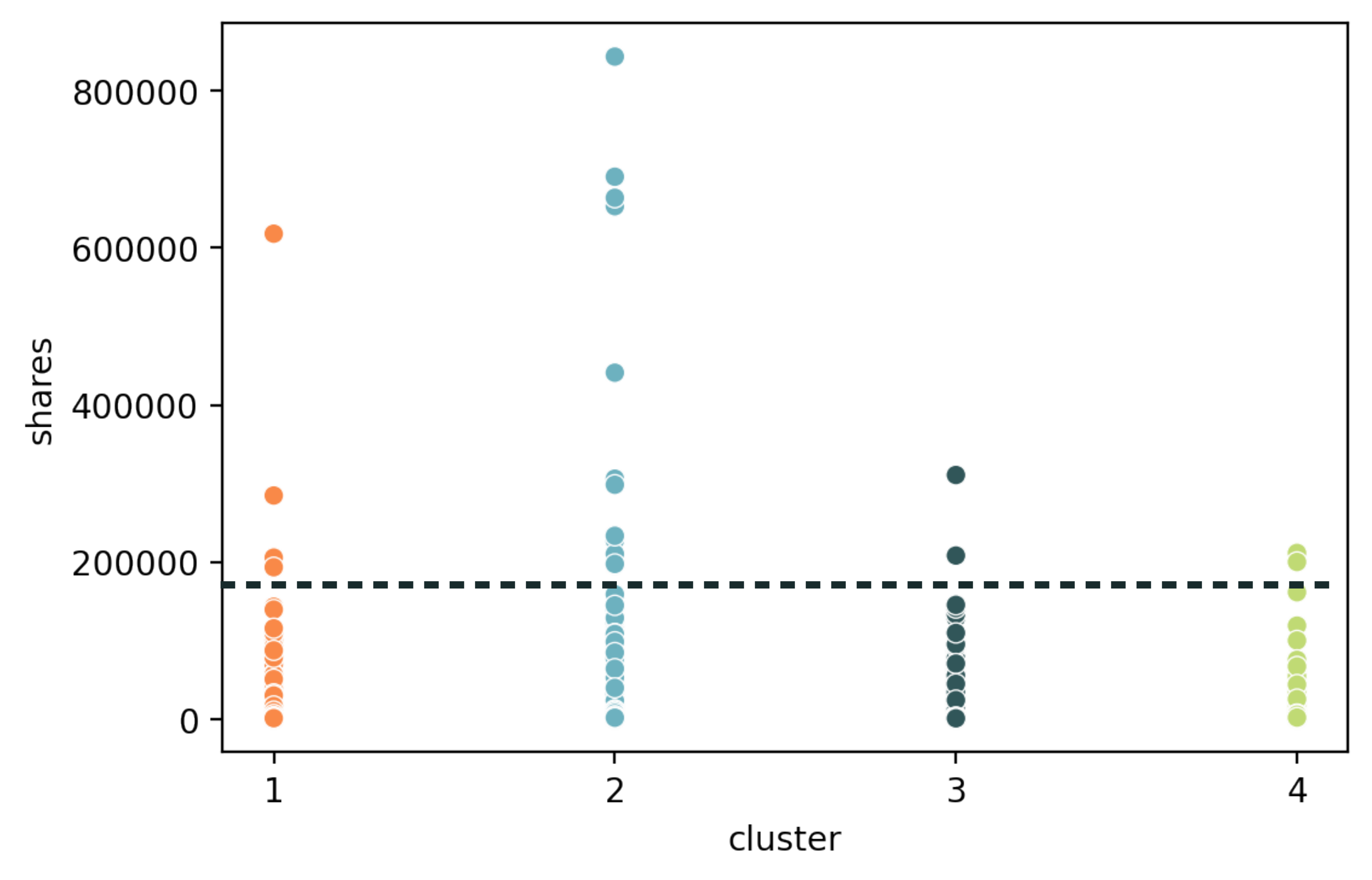
Cluster 4

- Articles with little to no text

Clustering

Cluster 2 had the most **popular** and **viral** articles followed by Cluster 1.

This could imply that the features present in Cluster 2, not only aid in popularity, but also to an article's likelihood of becoming viral.



Cluster 2

- Positive
- Subjective
- More text
- Short words

MODELING

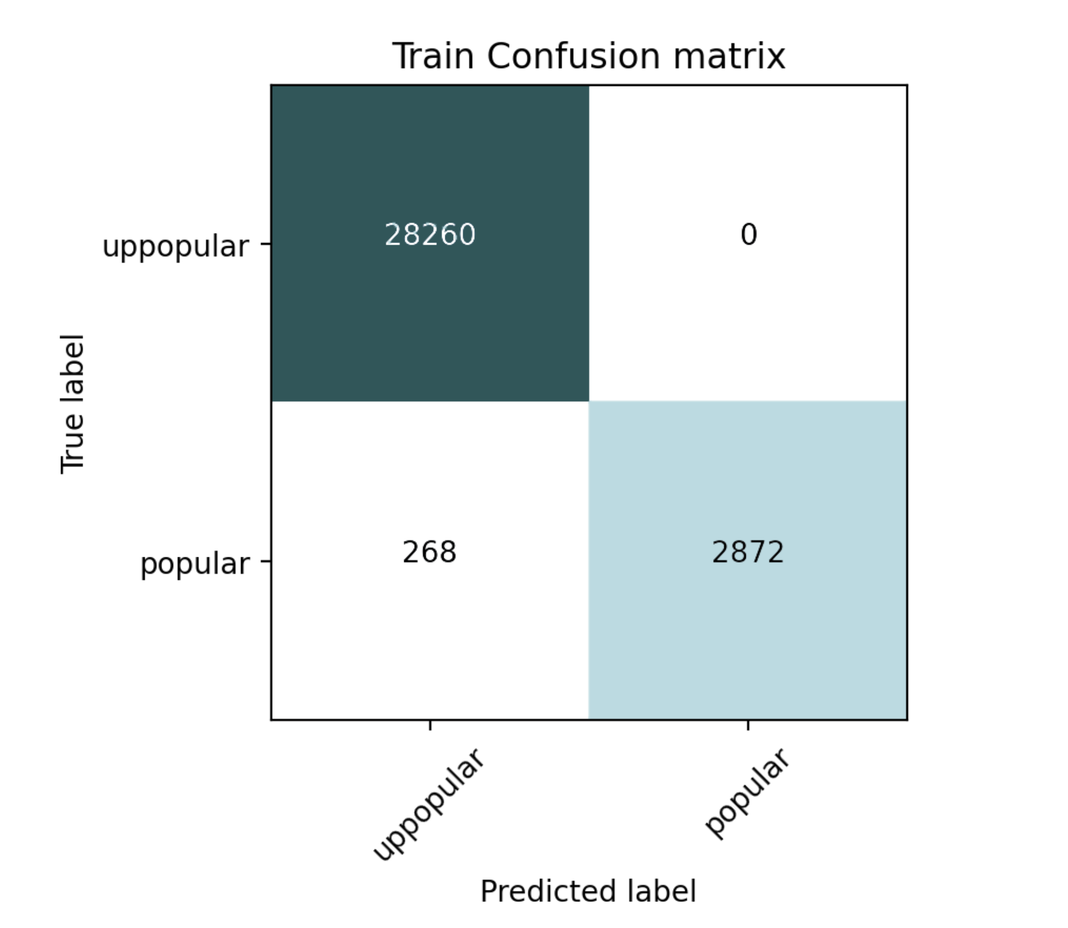
The Big Question

Can A Model Predict Popularity?

The Short Answer

Not Really…

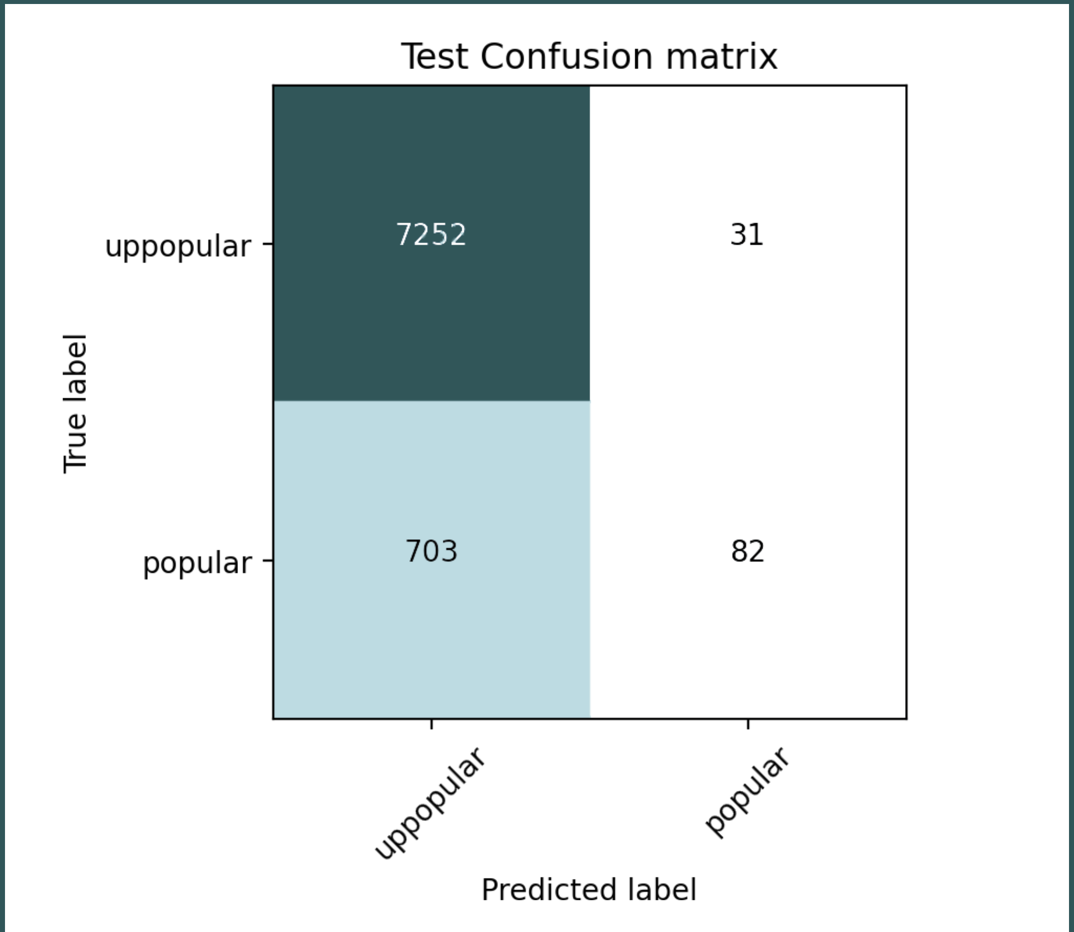
The Long Answer



Random Forest model for Binary Classification

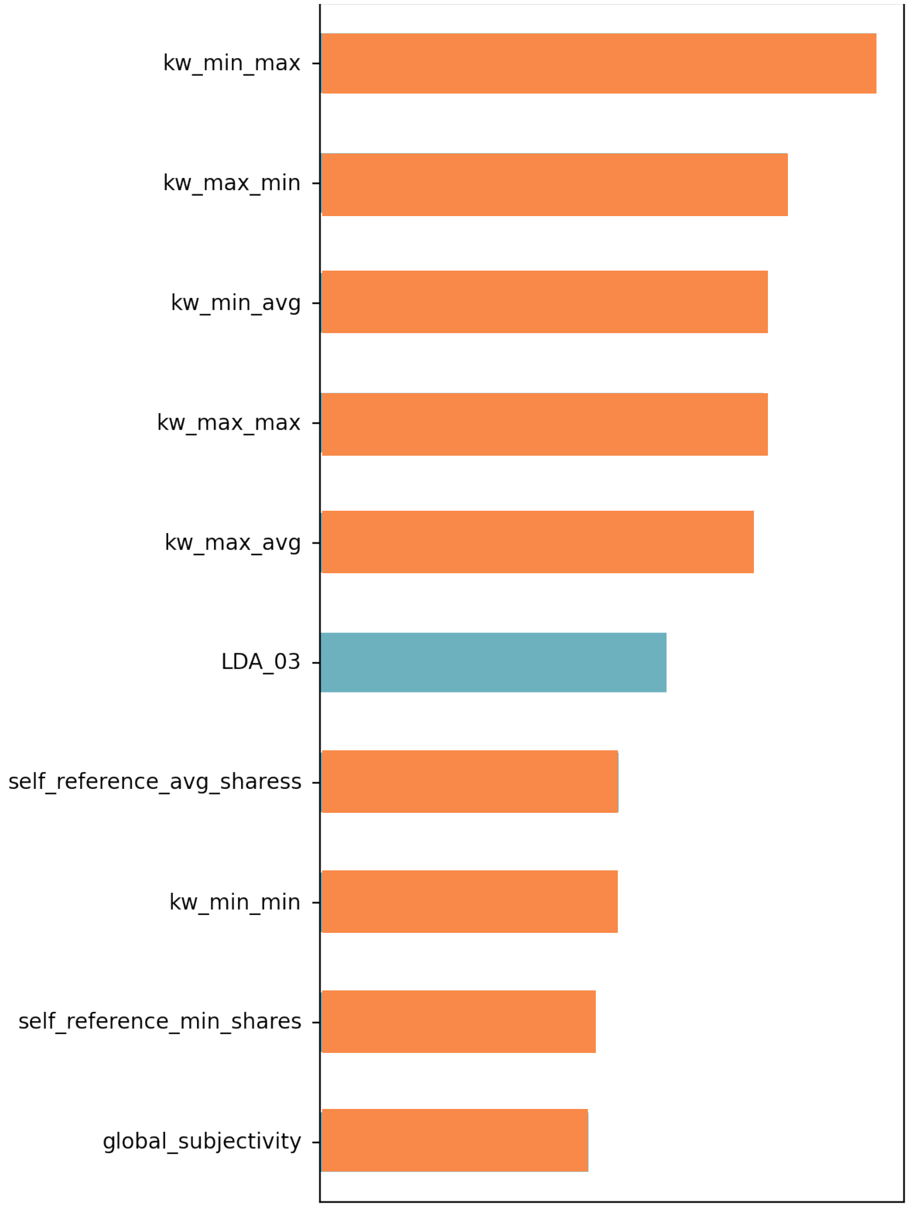
- Model could only identify a particular subset of popular articles.
- Even when testing the model with the training data, the model correctly labels only 91% of the popular articles

The Long Answer



- When tested on new data, the model could only identify 11% of the popular articles.
- Out of the articles it did label as popular, 73% were accurately labeled.

Overall, the model fails to identify all popular articles but does has a **73% accuracy** for those it did label as popular



Feature Importance

So what features are the most important factors in determining popularity according to the model?

- Shares of articles in the same keywords
- Shares of articles referenced
- Subjectivity

CONCLUSIONS

Putting it all together

So What Makes An Article Popular?

Text Features

Article Interconnectivity

Text Features

All readers want different things, but there are definitely some common traits among popular articles. Altogether the most popular articles tended to be:

- **Positive** — People like a **'feel good'** story
- **Long** — Don't be afraid of **a lot of text**
- **Simple** — Keep words **short** and **uncomplicated**
- **Subjective*** — Don't be afraid of being a little **subjective**

** Do try to avoid **subjective titles** though!*

Article Interconnectivity

Popular articles tend to attract attention to the articles related to it through keywords and references. Mashable could use this to their advantage by taking actions such as..

- Improve Suggested Articles
  - Continue to improve how articles are recommended to readers by **utilizing user viewing habits** or suggesting articles similar to what is currently being viewed.
  - Identify popular or viral articles early and provide these articles with more article recommendations.

Article Interconnectivity

Popular articles tend to attract attention to the articles related to it through keywords and references. Mashable could use this to their advantage by taking actions such as..

- Clean up the Keywords
  - There were 16,724 different keywords, only 8,035 (~48%) of which were used more than once.
  - Creating a **distinct set of keywords** for authors to choose from would help **build the connections** between articles.
    - Would also help the recommendation engine find similarities and create better suggestions