Leveraging Machine Learning for Predicting Customer Satisfaction

Ziad El Harairi, Juan Saavedra, Juan Ocampo

Methodology

Predictive Model

2 | Initial Model Exploration

3 | Dimensionality reduction

4 | Model Testing

1 | Exploratory Data Analysis

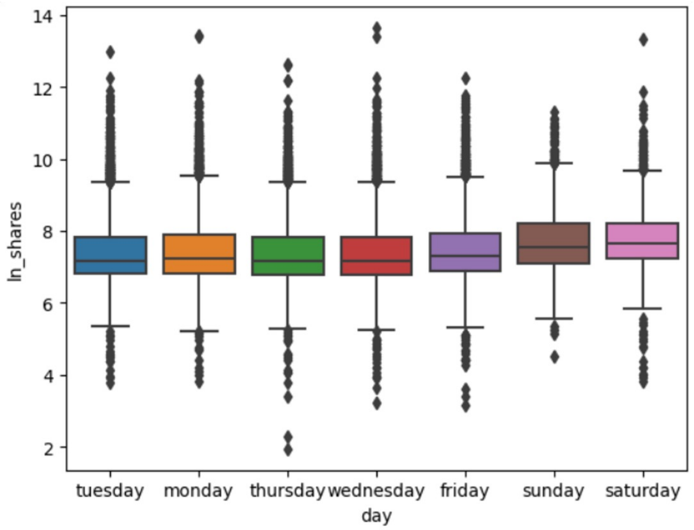
Business Case

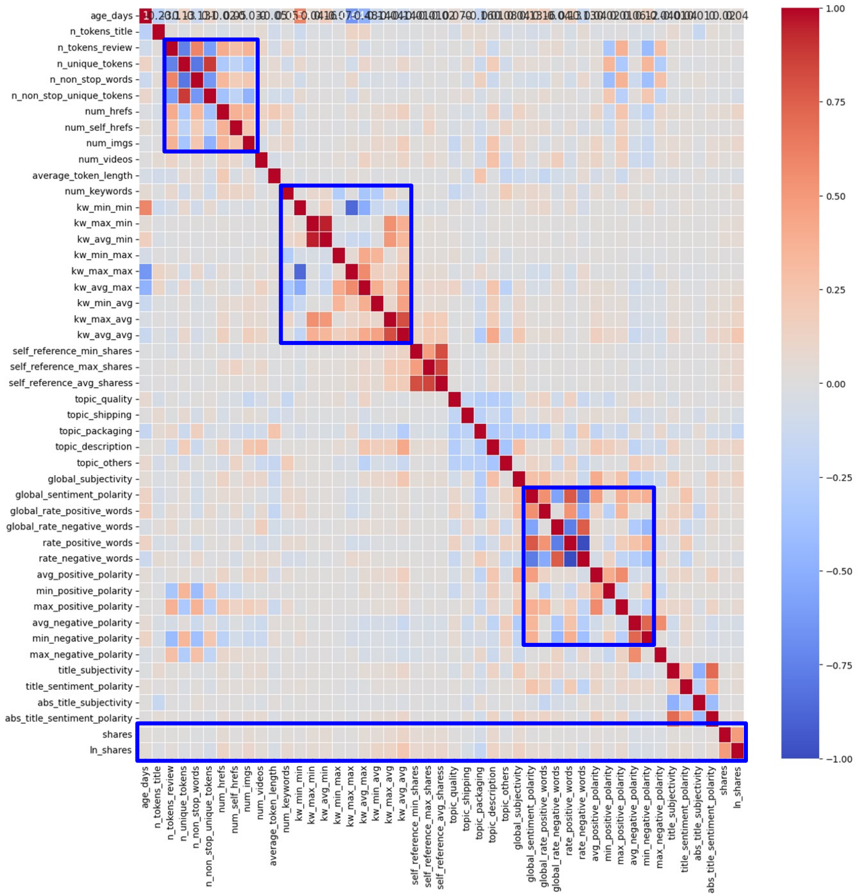
2 | Define what is a **Negative** review

3 | Define what is a **Popular** review

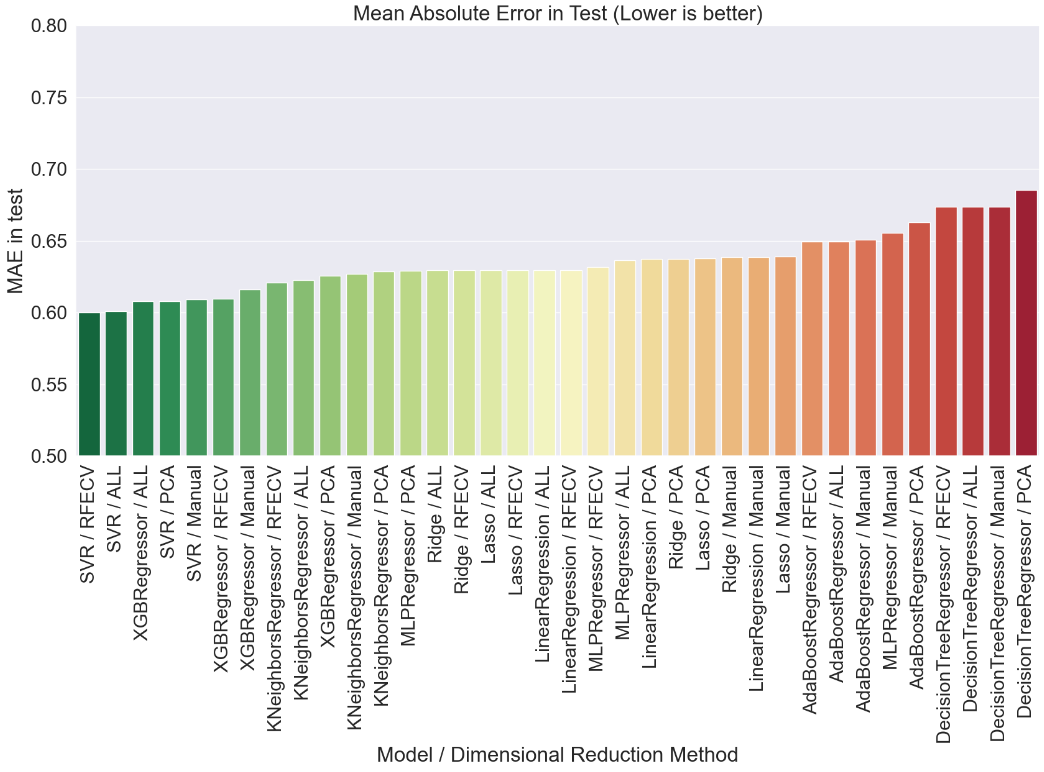
4 | Prediction Model

EDA


14
12
10
8
6
4
2
ln_shares
tuesday
monday
thursday
wednesday
friday
sunday
saturday
day


age_days
n_tokens_title
n_tokens_review
n_unique_tokens
n_non_stop_words
n_non_stop_unique_tokens
num_hrefs
num_self_hrefs
num_imgs
num_videos
average_token_length
num_keywords
kw_min_min
kw_max_min
kw_avg_min
kw_min_max
kw_max_max
kw_avg_max
kw_min_avg
kw_max_avg
kw_avg_avg
self_reference_min_shares
self_reference_max_shares
self_reference_avg_sharess
topic_quality
topic_shipping
topic_packaging
topic_description
topic_others
global_subjectivity
global_sentiment_polarity
global_rate_positive_words
global_rate_negative_words
rate_positive_words
rate_negative_words
avg_positive_polarity
min_positive_polarity
max_positive_polarity
avg_negative_polarity
min_negative_polarity
max_negative_polarity
title_subjectivity
title_sentiment_polarity
abs_title_subjectivity
abs_title_sentiment_polarity
shares
ln_shares
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Model Testing



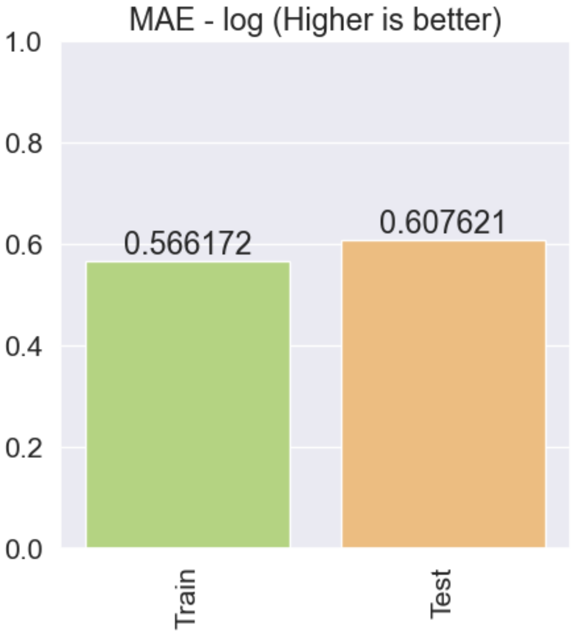
MAE TEST:

Best:
SVR/RFECV
0.59(2272)

Worst:
Decision Tree/PCA
0.68(2412)

Final Model

- XGBoost (Gradient Boosting)
- All predictors (45 numerical, 14 categorical dummies)



- MAE in tests 0.607 (2293)
- R2 in test 0.184

Business Case

Objective

Identify potentially popular negative reviews to prevent negative feedback

How

1 | Define what is a **Negative** review

2 | Define what is a **Popular** review

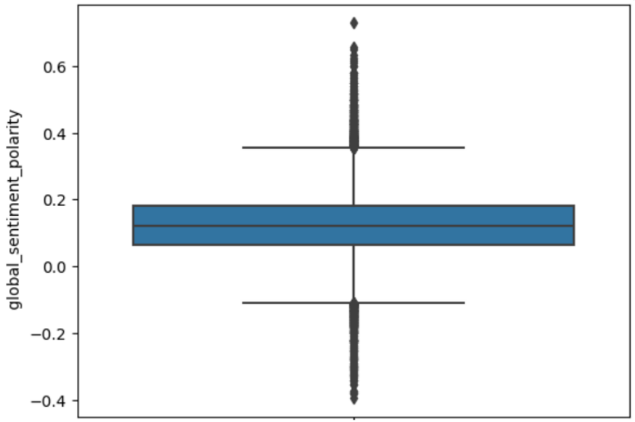
3 | Prediction Model

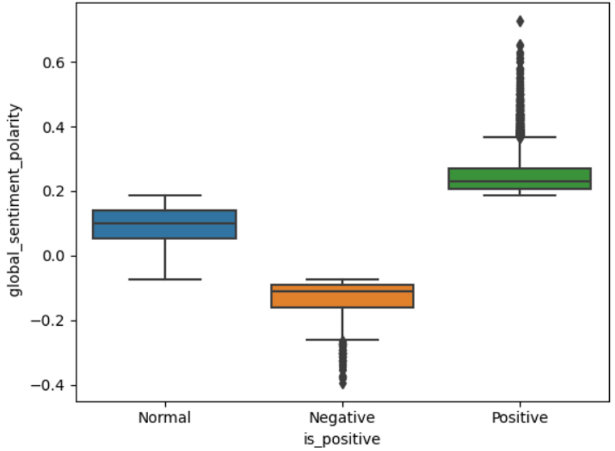
Business Case

1 | Define what is a **Negative** review

Rule

- Global sentiment polarity > mean + std -> Positive
- Global sentiment polarity < mean - std -> Negative

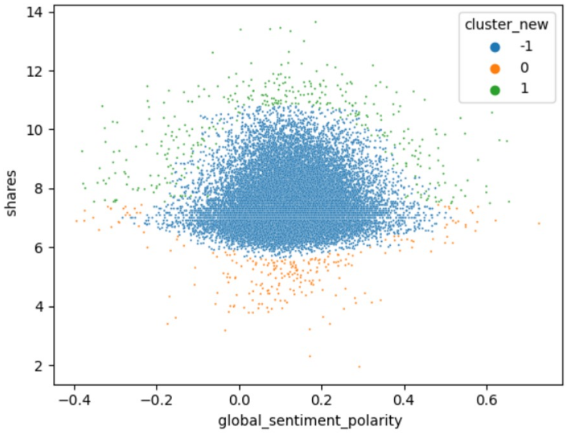




Business Case

2 | Define what is a **Popular** review

Hypothesis: Popularity hinges on sentiment polarity, with extreme values garnering less attention than moderate ones. Outliers indicate what's trending.



Cluster 1 - Interesting: Comments with a larger number of shares than usual for a given sentiment polarity

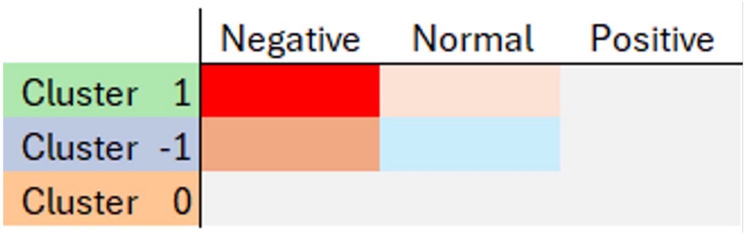
Cluster 0 - Not Interesting: Comments with a low amount of shares

Cluster -1 - Normal behaviour

Business Case

3 | Prediction Model

Classification Tree: Classify new comments to assign a level of urgency to be reviewed based on Clusters and Polarity



**98% accuracy

- Critical: Negative comment with potential to become popular
- High: Average negative comment
- Moderate: Neutral comment with potential to become popular
- Low: Average Neutral comment