EXPLORING SENTIMENT IN TWITTER TWEETS

CHLOE NGUYEN, ADVAYA GUPTA

MOTIVATION

POLARITY & SUBJECTIVITY

IN TWITTER TWEETS

THE DATA

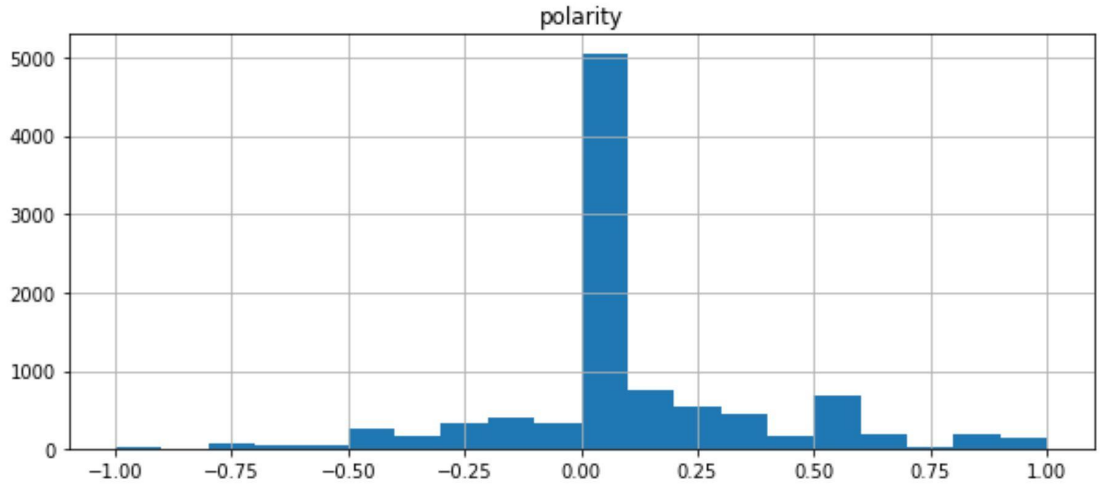
This dataset contains a mix of both popular and real time tweets (in extended mode) that include at least one hashtag sign and is in English .

There are 9999 observations and 4 features :

- **text:**: represents the tweet's text
- **favorite_count**: count of likes on a specific tweet
- **polarity**: ranges from -1 (negative) to 1 (positive)
- **subjectivity**: ranges from 0 (objective) to 1 (subjective)

polarity

Describes the tone of text. Ranges from -1 (negative) to 1 (positive).



Mean = 0.087396

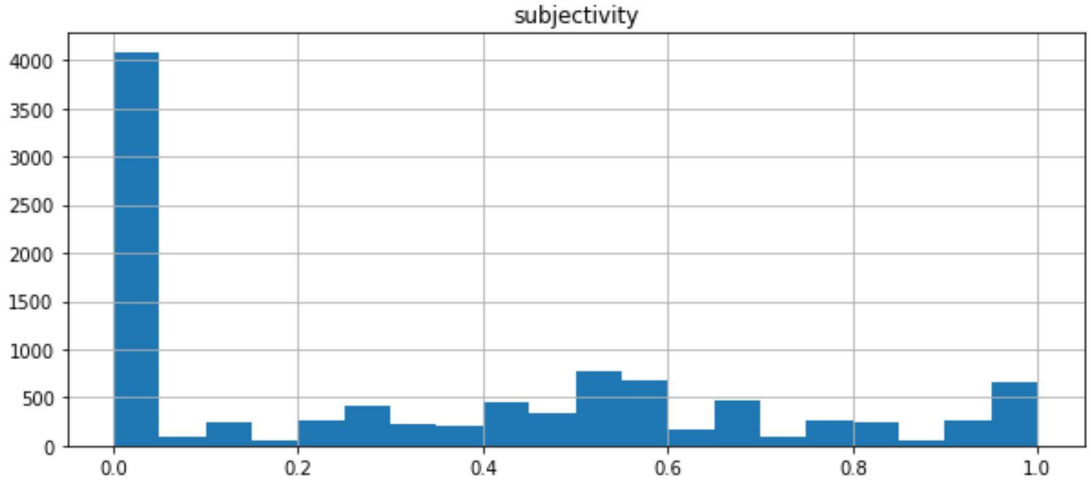
Median = 0

Min = -1.000000

Max = 1.000000

subjectivity

Describes how factual the text is. Ranges from 0 (objective) to 1 (subjective).



Mean = 0.332452

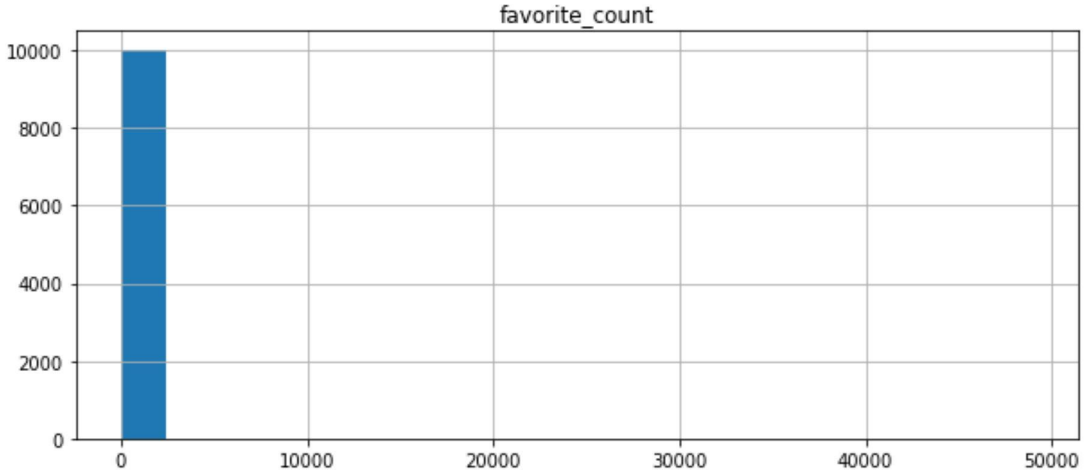
Median = 0.300000

Min = 0.000000

Max = 1.000000

favorite_count

Details the count of likes on a specific tweet



Number of tweets over 100 likes: 21

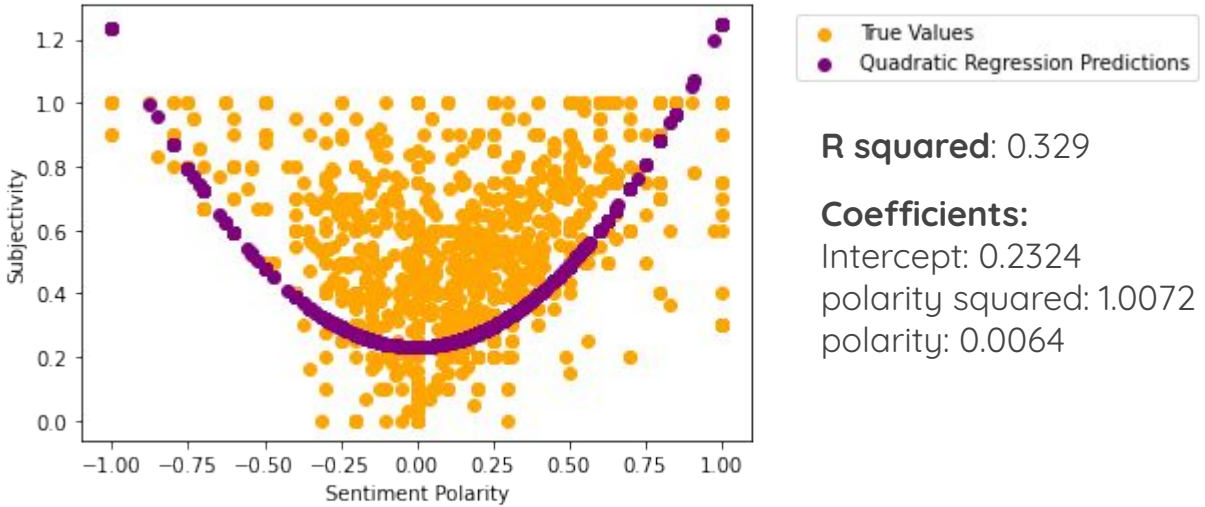
Number of tweets over 1000 likes: 3

Number of tweets over 10000 likes: 2

MODEL: QUADRATIC REGRESSION PREDICTIONS

Positive coefficient implies that the stronger the tone of the tweet, the more subjective the tweet would be.

Quadratic trend implies subjectivity increases faster for stronger tones.



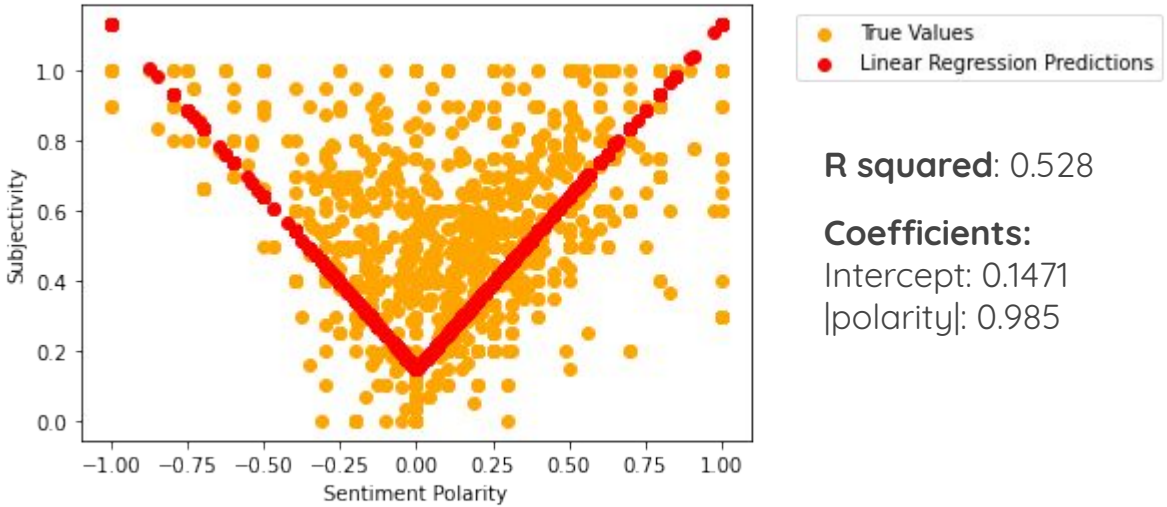
R squared: 0.329

Coefficients:
Intercept: 0.2324
polarity squared: 1.0072
polarity: 0.0064

MODEL: LINEAR REGRESSION PREDICTIONS

Again, the stronger the tone, the more subjective the tweet. In this case, rate of increase does not change.

Since the R squared for this model is higher, this is a better explanation for the relationship between polarity and subjectivity.



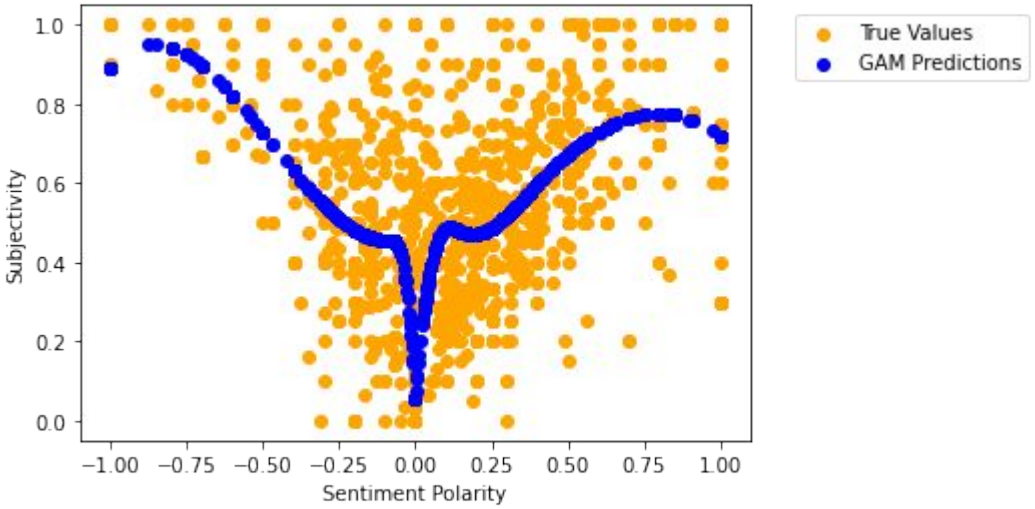
R squared: 0.528

Coefficients:
Intercept: 0.1471
|polarity|: 0.985

MODEL: GAM PREDICTIONS

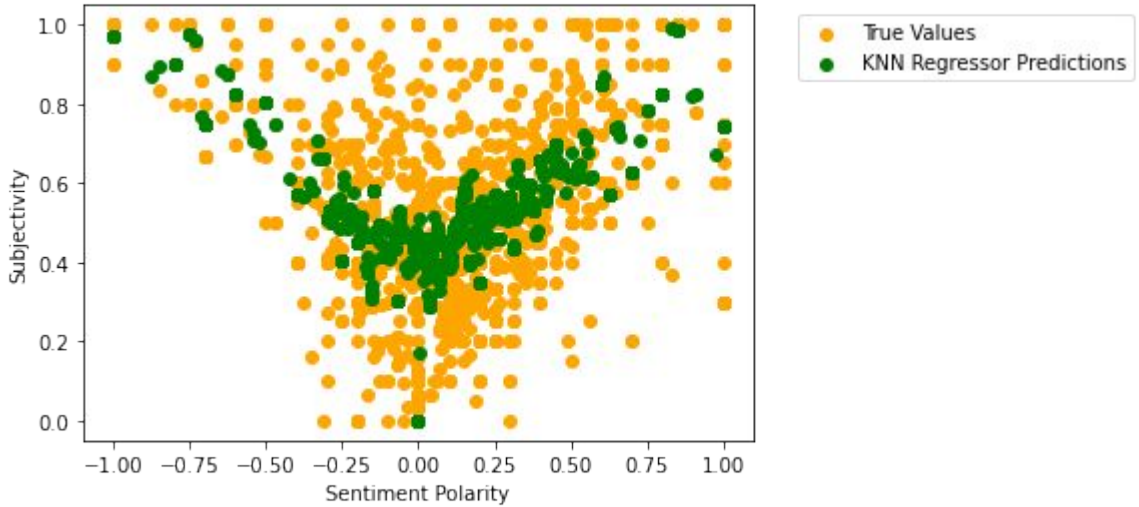
Estimates coefficient for multiple “basis” functions (that are added together) in order to fit the best possible curve.

Steep decline in predicted subjectivity for nearly neutral tweets.

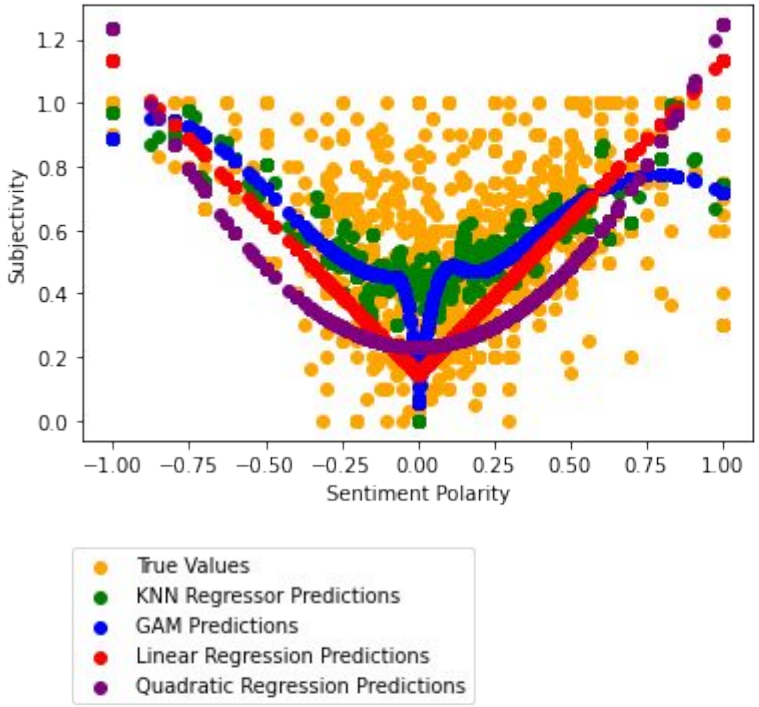


MODEL: KNN PREDICTIONS

Averages the subjectivity for the 17 nearest sentiment values from the training set.



RESULTS



Model	RMSE
Quadratic Regression	0.2808
Linear Regression	0.2343
GAM	0.1957
KNN	**0.0370**

CONCLUSION:

Our results imply that subjectivity has a positive association with strength of tone (polarity).

THANK YOU

That concludes the presentation.