How do Gamers Feel?

Joshua Sonnen
sonnenj@wwu.edu

Original Data

Mohamad Tarek performed a web scraping of Steam to pull user review data. There were a total of 762,851 reviews across 290 games.

Review Features

- ★ Recommendation (binary)
- ★ Hours Played
- ★ Funny votes
- ★ Helpful votes

Game Features

- ★ Game_name (290)
- ★ Publisher (173)
- ★ Developer (216)
- ★ overall_player_rating (11)

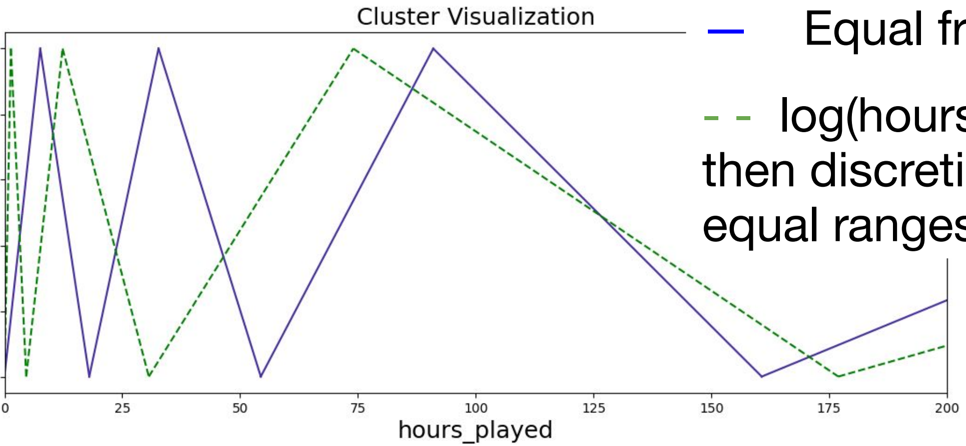
TextBlob Generated

TextBlob is a Python NLP library. I use its sentiment analysis tool to generate additional features.

Review Features

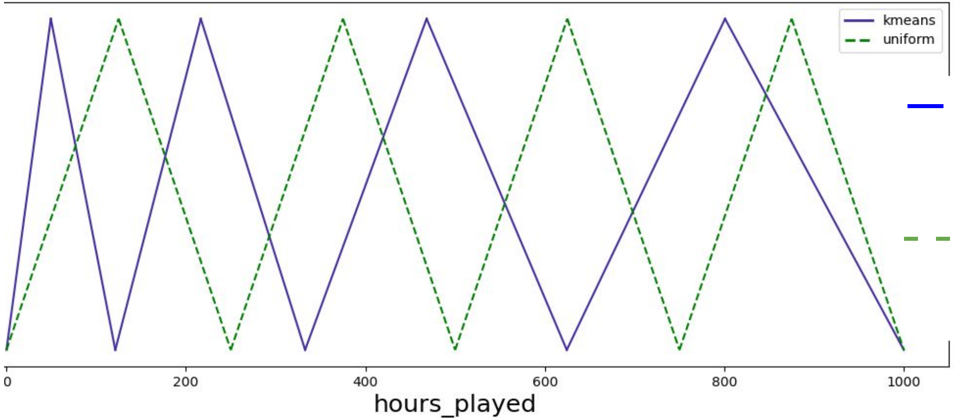
- ★ Number of words
- ★ Number of sentences
- ★ Polarity [-1, 1]
- ★ Subjectivity [0, 1]
  –"How opinionated the review is"

Hours Played Discretization Strategies



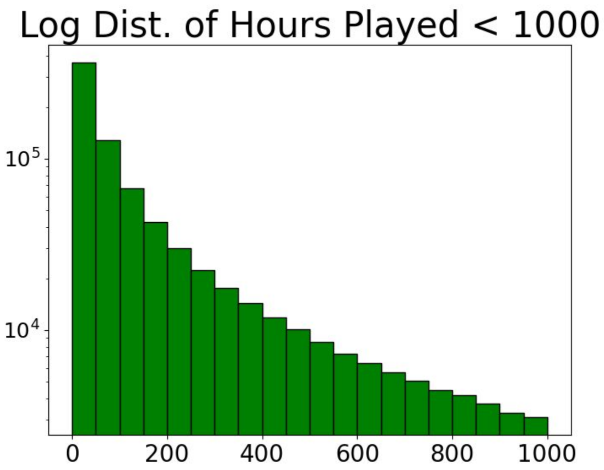
— Equal frequency

- - log(hours_played) then discretize by equal ranges

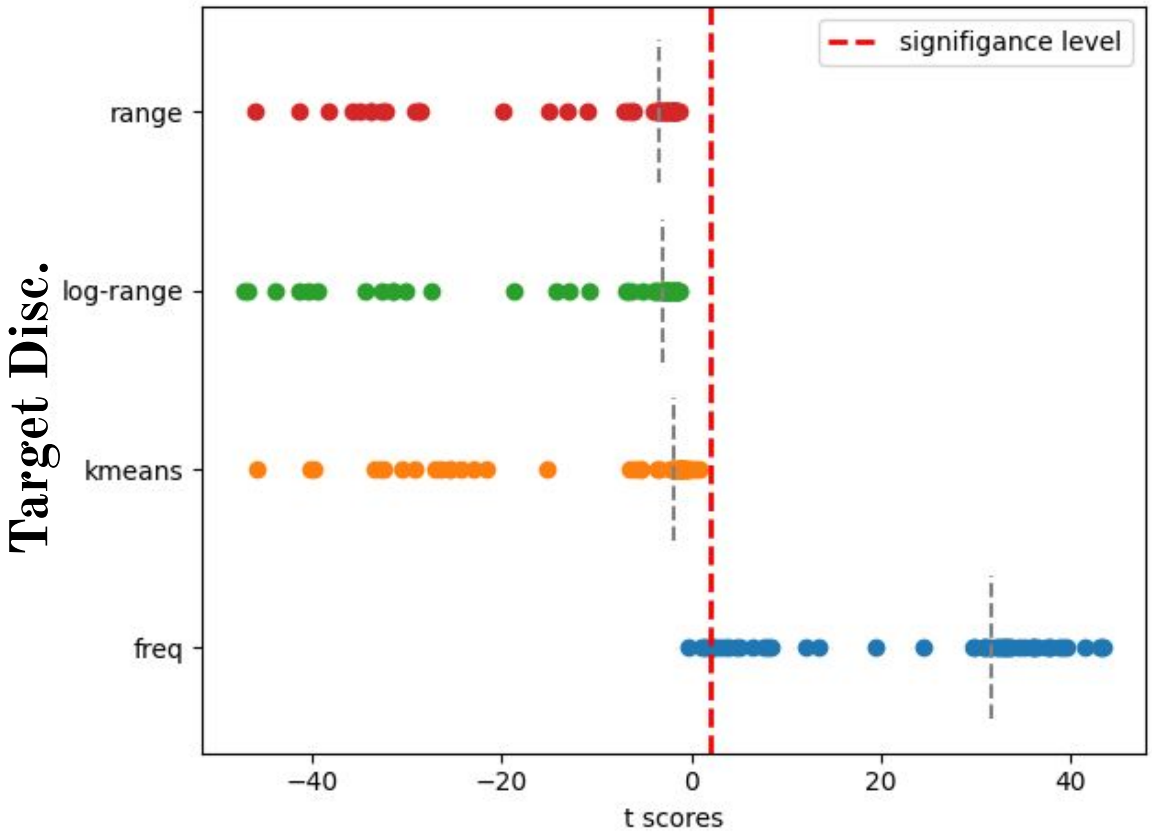


— K Means Clustering

- - Equal ranges



Target Discretization Analysis



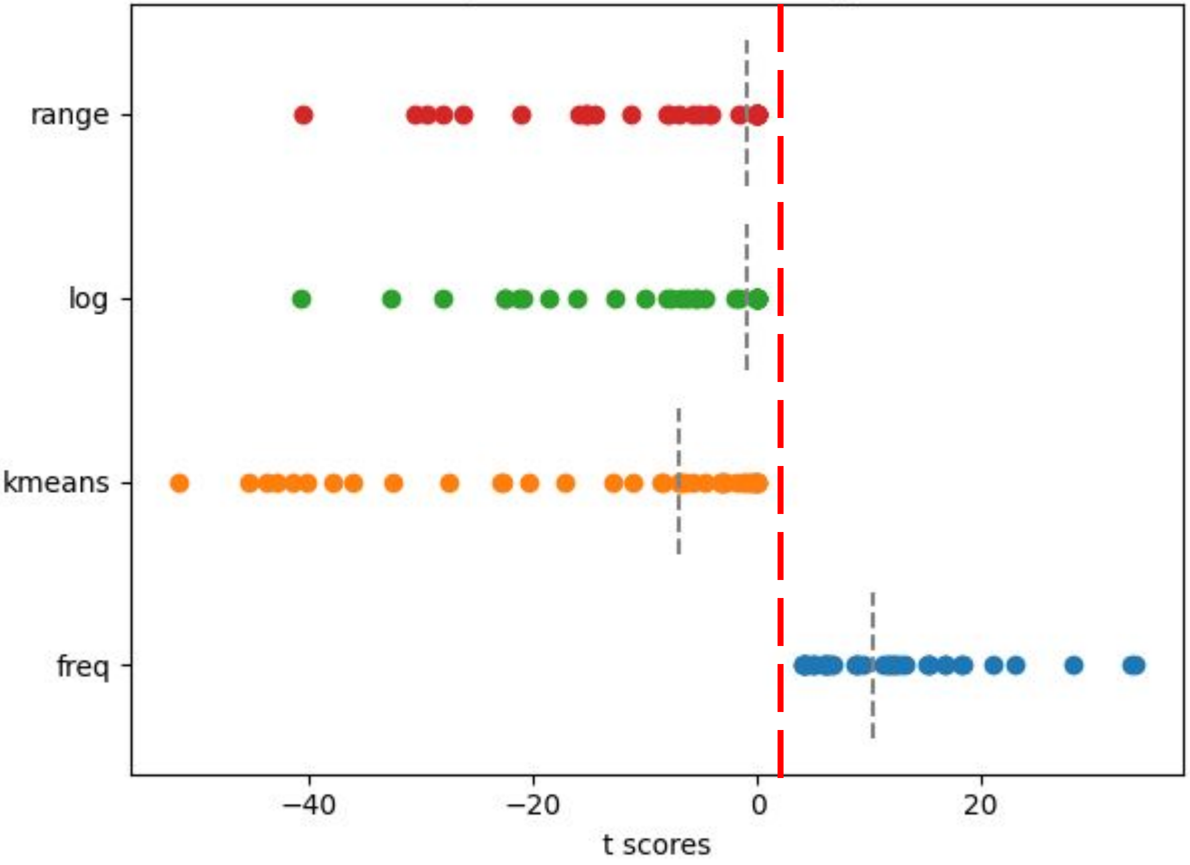
- Random Forest classifier.
- Bootstrap factor of 3
- 11 feature subsets
- 4 split criteria: random, entropy, gini, log loss
- 4 discretization strategies: range, log-range, k-means, and frequency

Each t-score represents the difference in sample means b/w the majority classifier and the random forest. I use Welch's t-test to correct for unequal population variance.

$T_{\text{critical}} \approx 2.04$

Significance threshold

Target Discretization Analysis



- K Neighbors classifier
- Bootstrap factor of 3
- 5 feature subsets
- 4 bin sizes
- 2 choices of k

Each t-score represents the difference in sample means b/w the majority classifier and the classifier. I use Welch's t-test to correct for unequal population variance.

$T_{critical} \approx 2.04$
Significance threshold

Bootstrapping

With 760,000 data points, my first step is to remove some data

1. Sample with replacement. These indices are dropped
2. Set difference → Return not sampled indices
3. 80% train; 20% validation split

Sampling k points on the range (0, N) I will expect

$$\text{\# of unsampled data} = e^{-k/n}$$

$$n = 762{,}851$$

$$k = 1n : (e^{-1}) = 0.3678n \rightarrow 280{,}637$$

$$k = 3n : (e^{-3}) = 0.0498n \rightarrow 37{,}990$$

$$k = 5n : (e^{-5}) = 0.00674n \rightarrow 5{,}119$$

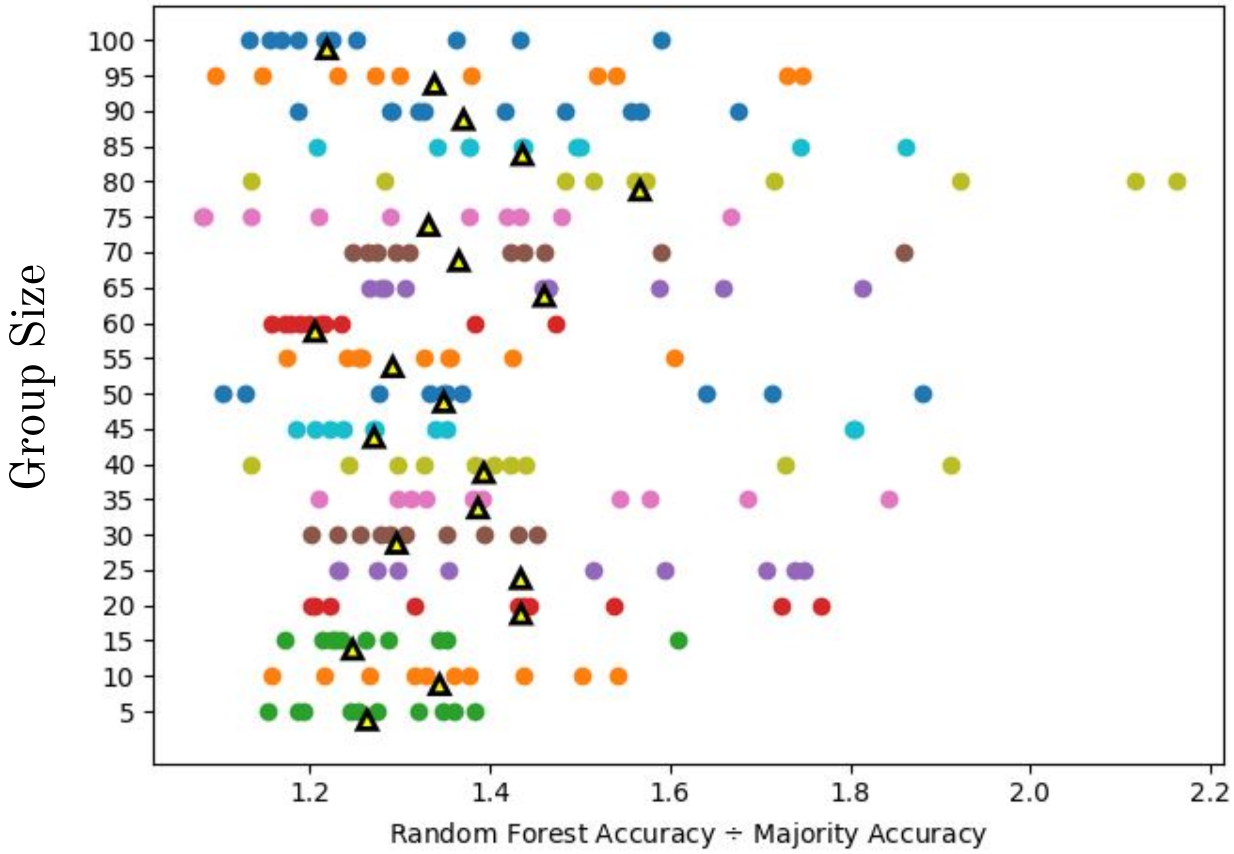
For K Nearest Neighbors classification, I use a bootstrap factor of 3.

For Random Forest classification, I use a bootstrap factor of 5.

Bin Sizes

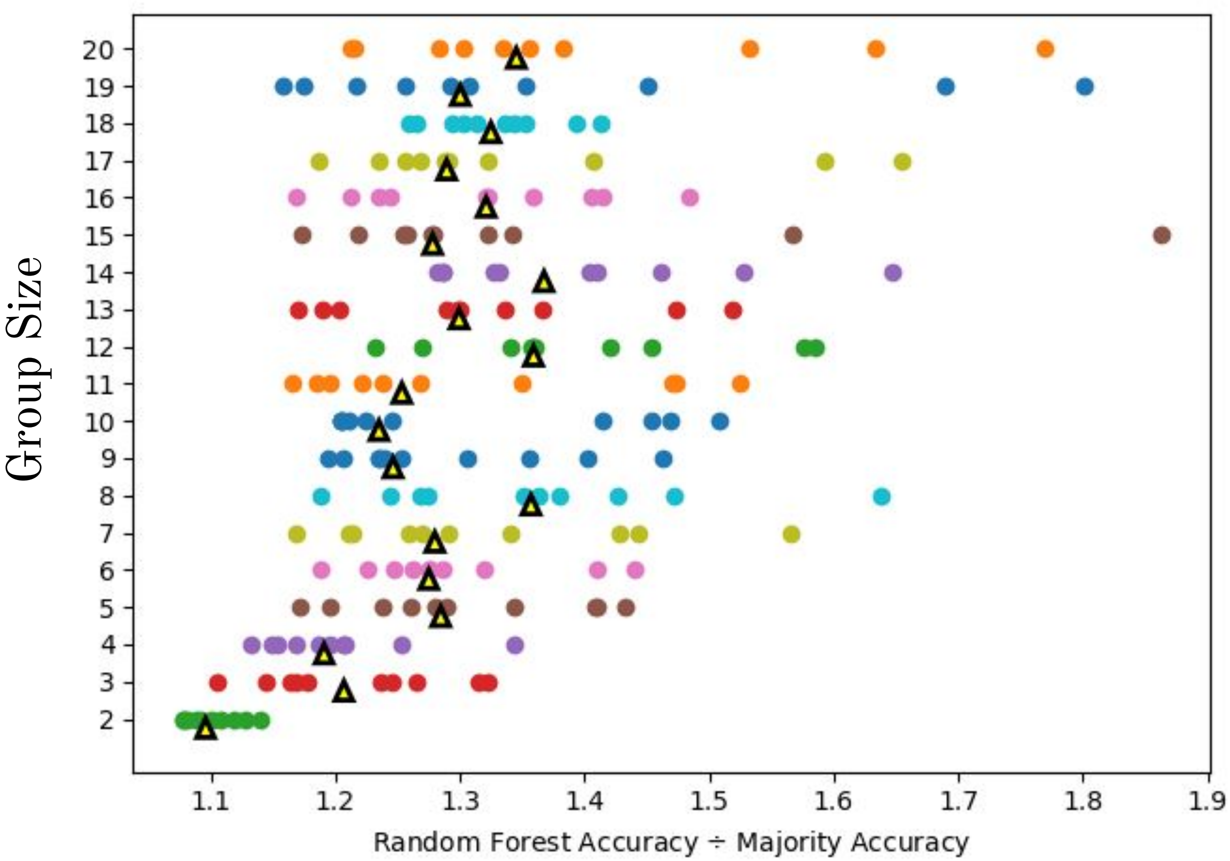
- Random Forest
- Bootstrap factor of 3
- All features
- Random split criterion
- Max depth = 5
- # Trees = 10

Classification becomes more difficult with more categories. My goal is to find a bin size which hours played will naturally cluster into.



- Random Forest
- Bootstrap factor of 3
- All features
- Random split criterion
- Max depth = 5
- # Trees = 10

Detailed look at the first 20 group sizes. I observe a fall-off in median classifier performance at 8.



2 Clustering– Motivated Clustering Bin Sizes

For each continuous attribute:

    For each category:

        Cluster into 8 groups

        Calculate Silhouette score

        Cluster each subgroup into 2 groups using k means clustering

        Evaluate cluster using silhouette score

Continuous variable	Hours played	Helpful votes	Funny votes	Number of words	Number of sentences	Polarity	Subjectivity
Group Size	9	11	10	9	10	11	11

Feature Subsets

Random Forest Parameters

Nine Hopeful Subsets

1. game_name, overall_player_rating
2. publisher, overall_player_rating
3. developer, overall_player_rating
4. game_name, publisher, developer, overall_player_rating
5. recommendation/hours played, game_name, overall_player_rating
6. recommendation/hours played, publisher, overall_player_rating
7. recommendation/hours played, developer, overall_player_rating
8. recommendation/hours played, game_name, publisher, developer, overall_player_rating
9. All categorical AND continuous features.

Each trial is ran 30 times to provide enough samples for t-testing

Bootstrap factor of 5.

Ensemble 100 trees
Max Features = # features
Bin size = 8

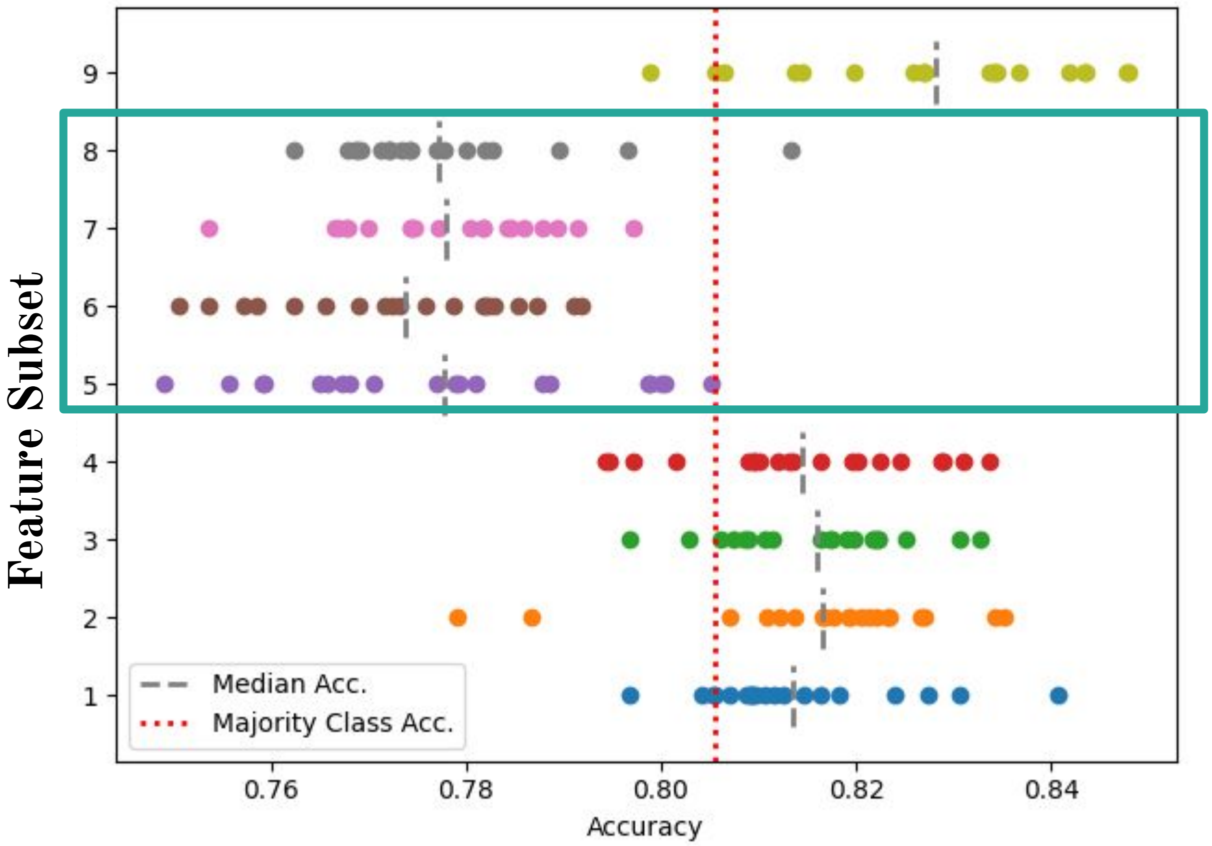
Criteria

- Entropy
- Gini Index
- Log loss
- Random

Target Disc

1. Uniform
2. K means
3. Frequency
4. Equal range on log(hours_played)

Predict Recommend– Feature Subset Analysis

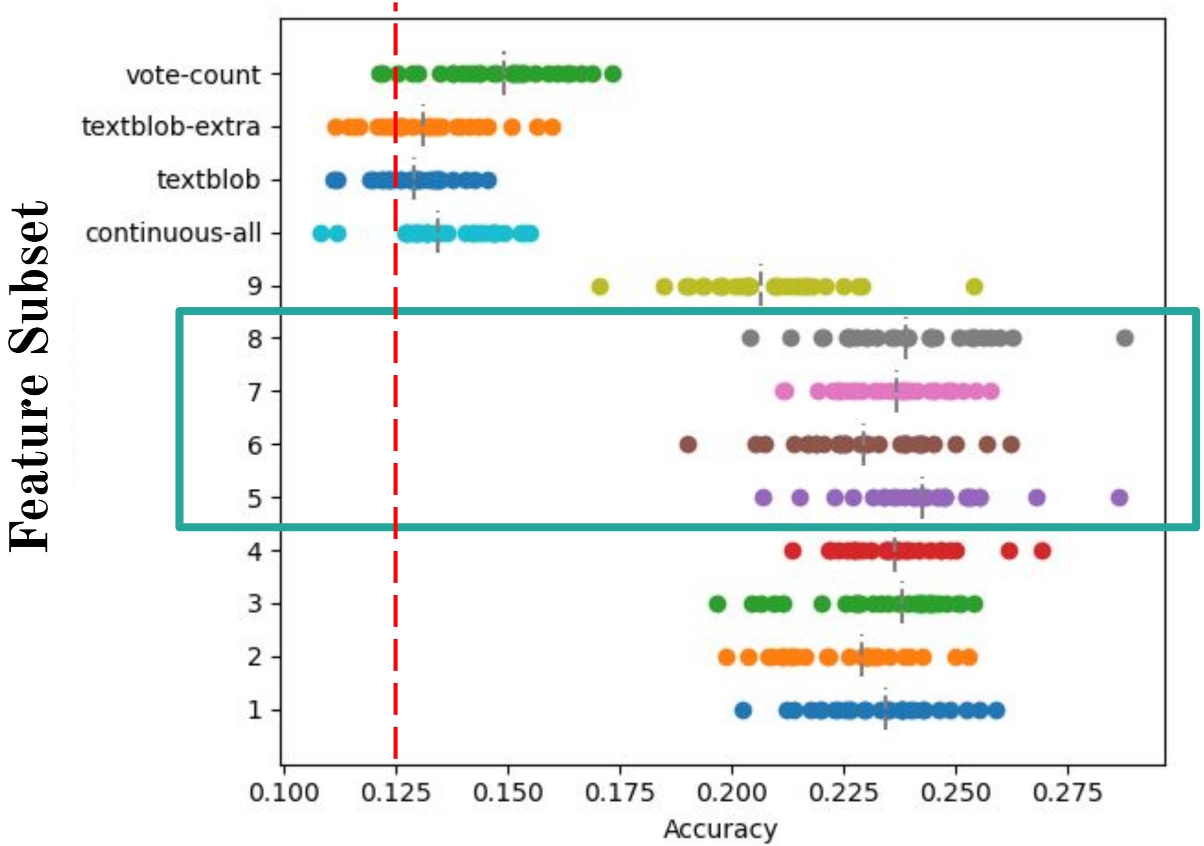


– Random Forest
– Random split criterion

Feature subsets 5-8 are the same as 1-4 except they also have hours played as a feature. It is unusual for the addition of a feature to hinder prediction capabilities.

Hours played is a continuous feature. Perhaps the continuity makes it difficult to split.

Predict Hours Played– Feature Subset Analysis



Recommendation does not have an influence on the classifying performance for predicting hours played.

The 4 worst performing feature groups are all continuous features, supporting my hypothesis that continuous features hinder Random Forest classification.

TextBlob Features

Trial 1: TextBlob Features

<u>TextBlob Groups</u>

1. TextBlob 1 – 'num_words'
2. TextBlob 2 – 'num_sentences'
3. TextBlob 3 – 'polarity'
4. TextBlob 4 – 'subjectivity'
5. TextBlob_length – 'num_words', 'num_sentences'
6. TextBlob_sentiment – 'polarity', 'subjectivity'
7. TextBlob_pair1 – 'num_words', 'polarity'
8. TextBlob_pair2 – 'num_words', 'subjectivity'
9. TextBlob_pair3 – 'num_sentences', 'polarity'
10. TextBlob_pair4 – 'num_sentences', 'subjectivity'
11. TextBlob_all – 'polarity', 'subjectivity', 'num_words', 'num_sentences'

Bootstrap factor of 5.

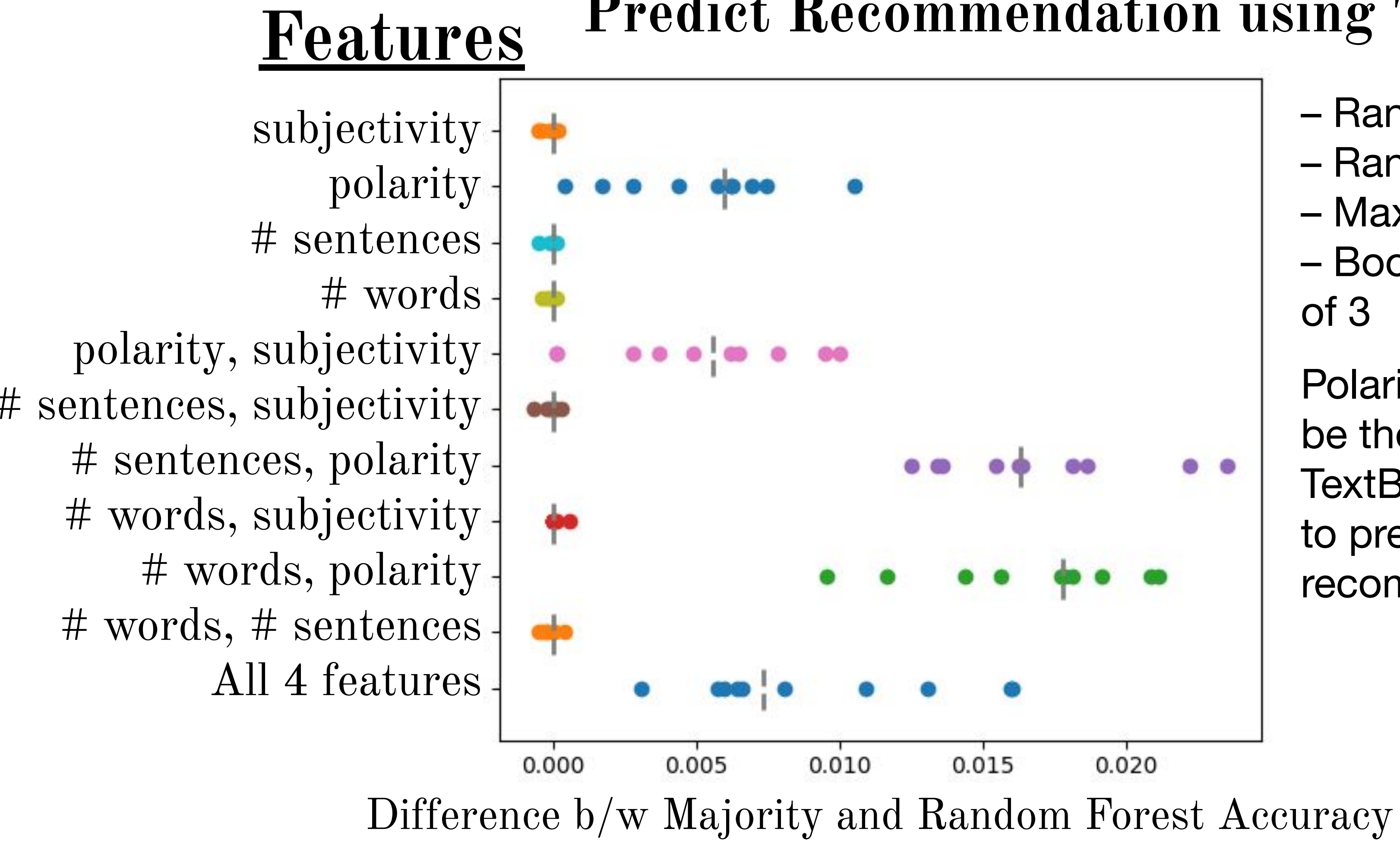
Each trial is ran 30 times to provide enough samples for t-testing

Ensemble 10 trees
Max Features = # features

Criteria

- Entropy
- Gini
- Log
- Random

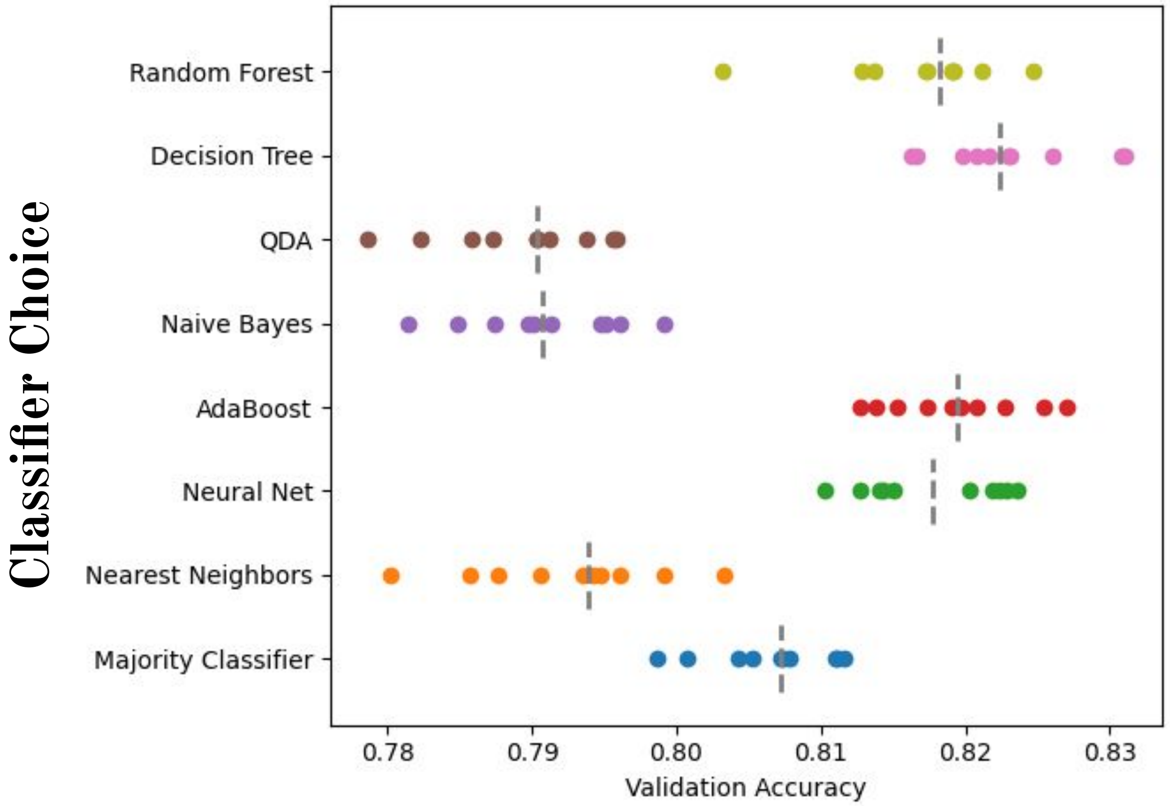
Predict Recommendation using TextBlob



- Random Forest
- Random Split
- Max Depth = 5
- Bootstrap factor of 3

Polarity seems to be the best TextBlob feature to predict recommendation.

Predict Recommend– TextBlob Classifier Analysis

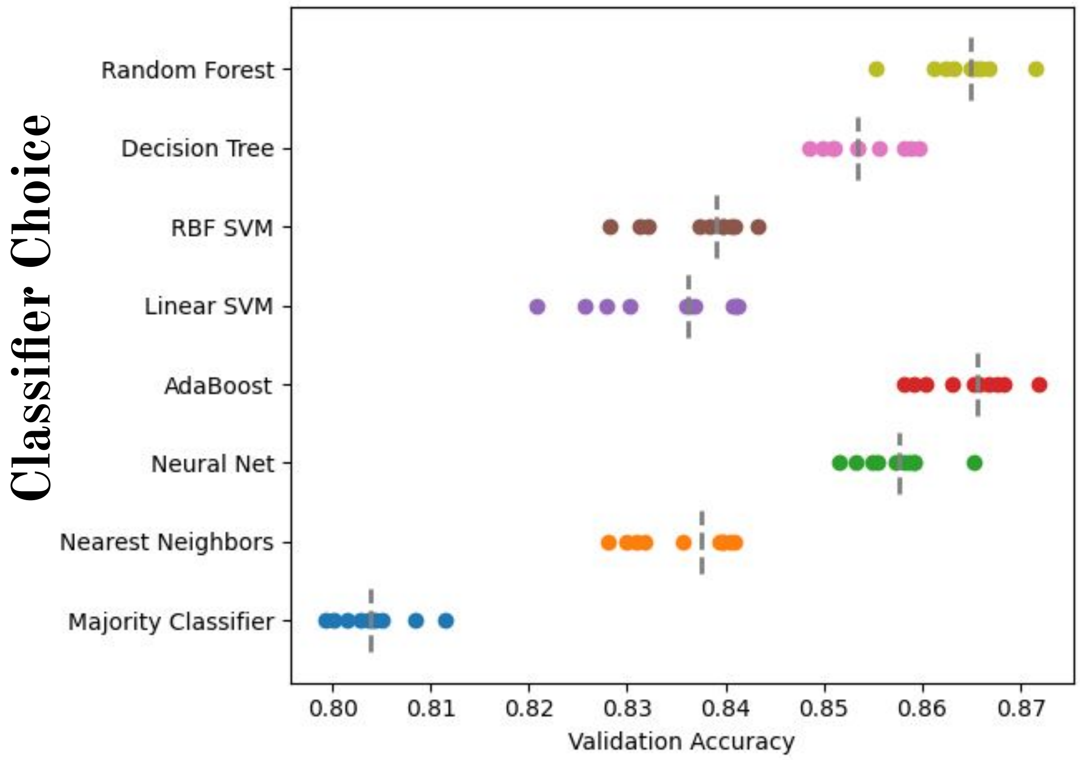


– # words, # sentences, polarity, subjectivity

These classifiers do not achieve state-of-the-art performance. Later, you will see accuracy up to 87% using all features.

All Classifiers

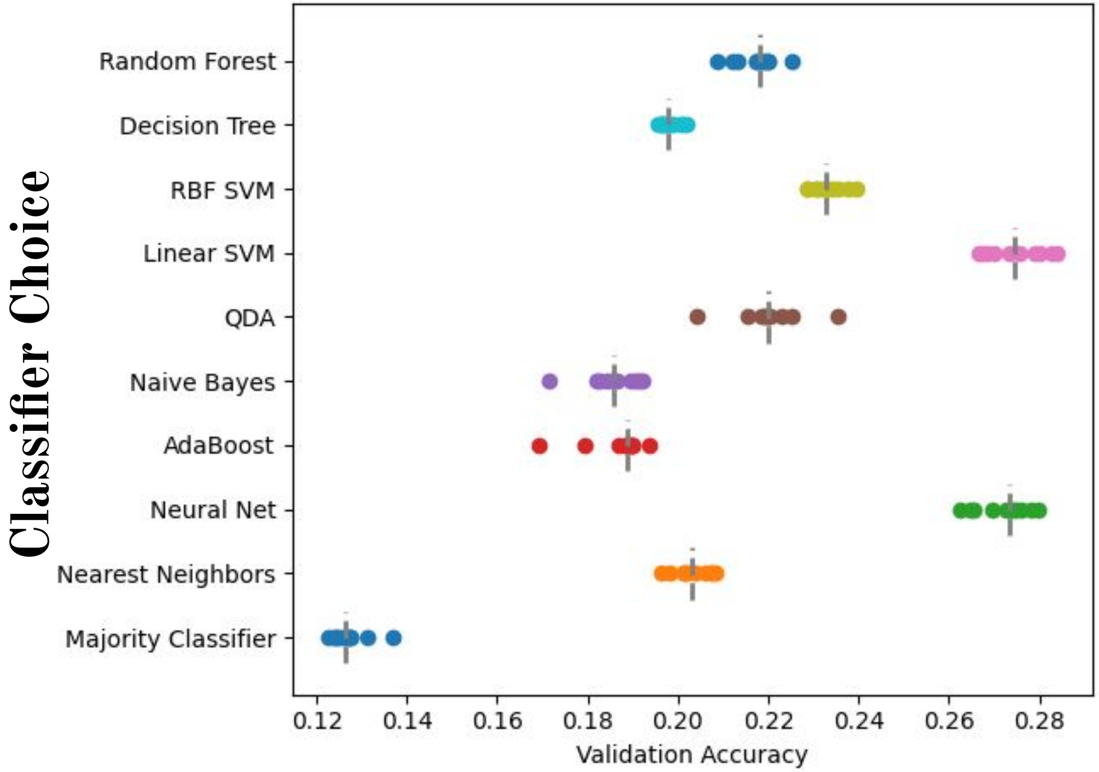
Predict Recommend– Classifier Analysis



– All features
– Bootstrap factor of 3.
– Sensible classifier parameters.

QDA and Naive Bayes performed below 60% accuracy.

Predict Hours Played– Classifier Analysis



– All features
– Bootstrap factor of 3.
– Sensible classifier parameters.

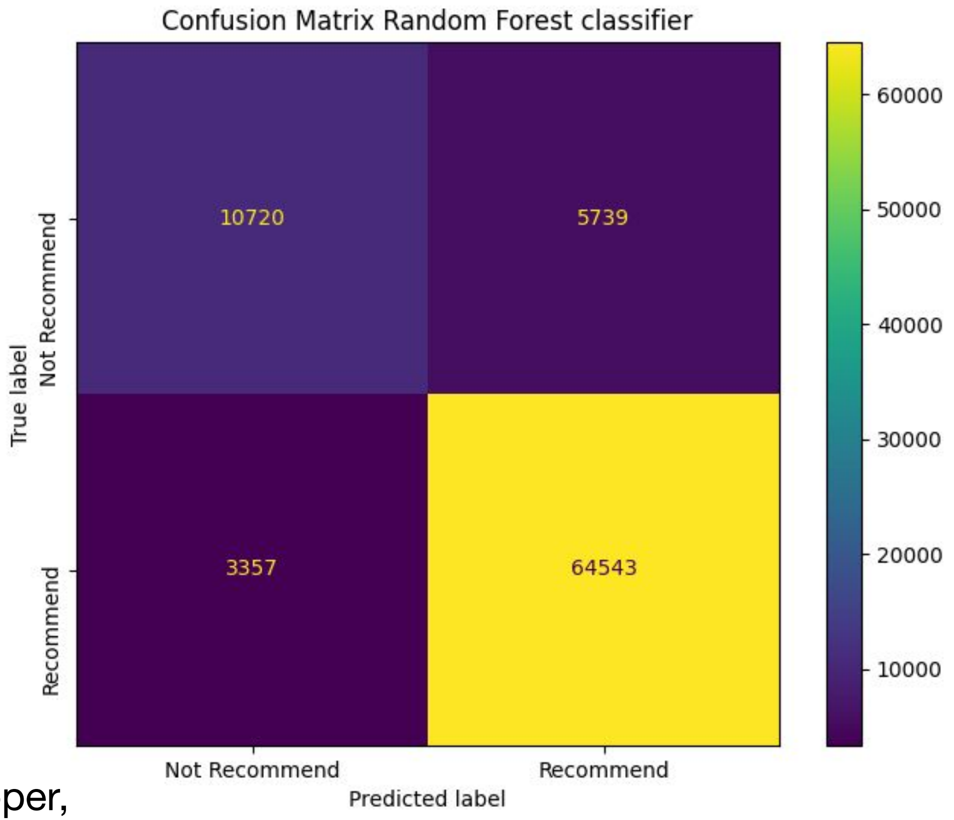
All classifiers improved over the Majority.

Test Set Evaluation

Test Set Evaluation
Predict Recommendation

Classifier Accuracy– 89.22%
Majority Class Accuracy– 80.49%
8.73% Improvement

Random Forest
Split Criterion = Random
Trees = 100
Feature Subset = ALL
hours_played, game_name, publisher, developer, overall_player_rating, Helpful, funny, num_words, num_sentences, polarity, subjectivity

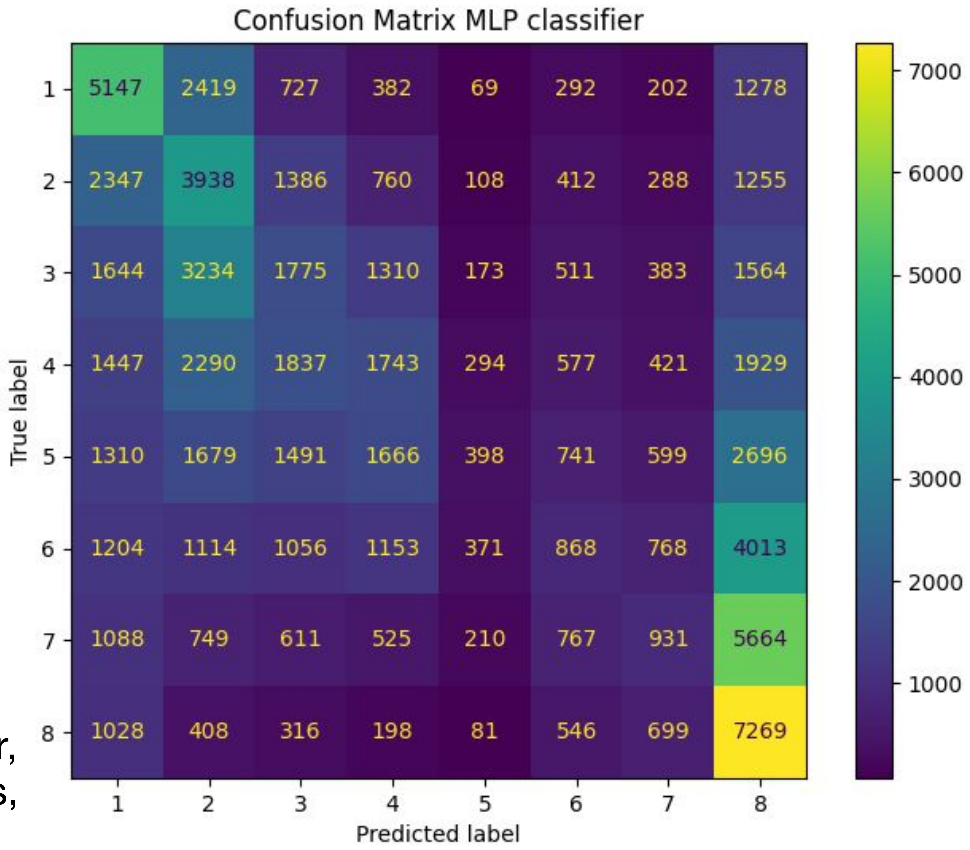


Test Set Evaluation
Predict Hours Played

Classifier Accuracy– 26.16%
Majority Class Accuracy– 12.49%
13.67% Improvement

Neural Network– MLP Classifier
- Alpha = 1 ; max_iter = 1000

Discretization Strategy = equal frequencies
Group Size = 8
Feature Subset = ALL
hours_played, game_name, publisher, developer, overall_player_rating, Helpful, funny, num_words, num_sentences, polarity, subjectivity



Future Expansions

- I only analyzed feature subsets for Random Forest and K Nearest Neighbors. Expand search to other classifiers.

- Group size of 8 is an arbitrary choice. Anywhere in the range [2, 40] could be an acceptable choice for group size. Or use a model capable of regression.

- Utilize Associative Rule Mining techniques to find optimal feature subsets. I brute forced a subset of feature groupings, but this only extended to TextBlob combinations (excluding triplets) and 9 arbitrary groupings.