

Text Sentiment Analysis Prediction Using Disneyland Reviews

David Berberena

Bellevue University

DSC 680 Applied Data Science

Amirfarrokh Iranitalab

November 17, 2024

Text Sentiment Analysis Prediction Using Disneyland Reviews Draft Paper

Business Problem

The Walt Disney Company has grown exponentially from its humble start as an animation studio to what may very well be the most common household name in the United States regarding the entertainment industry. The Company comprises various parts that create its economic fingerprint: Parks and Resorts, Media Networks, Studio Entertainment, and Consumer Products. The Parks and Resorts sector of the Company has been facing recent customer retention issues due to many factors and business changes, so understanding the guests' sentiments regarding their experience at Disneyland parks across the globe better positions the Company to be proactive rather than reactive to guest commentary. Review sentiment can typically be classified as either positive, neutral, or negative, and the automation of such classification to drive decision-making geared towards the improvement of the guest experience will help the Walt Disney Company abate the recently waning attendance at their theme parks. Predicting customer sentiment through their reviews will also strengthen the relationship between the Company and the guest.

Background/History

The large corporation the world knows as the Walt Disney Company, which holds many entertainment offerings such as Pixar, ABC, ESPN, Marvel Entertainment, Lucasfilm, FOX, Hulu, and more, began in 1923 as Disney Brothers Cartoon Studio, founded by Walt and Roy Disney (Tenebruso, 2020). In 1928, *Steamboat Willie* came to be, introducing one of the most beloved animated characters of all time, and in 1937, the cartoon studio's *Snow White* film was named

the highest-grossing movie of that time. In 1955, Walt and Roy Disney opened their first theme park in Anaheim, CA, and called it Disneyland. This legacy has grown to include the successful operation of six theme park resorts encompassing more than fifteen theme parks across three continents, many new franchises, and millions of guests enjoying Walt Disney's dreams made a reality. These guests who vacation at Disney parks across the globe express their sentiments of gratitude, happiness, and fun, as well as annoyance, disbelief, and disappointment in the effort to have the Company better operate the theme parks. These sentiments in the form of reviews are one of the most direct ways the Company can access how guests think and feel about the overall Disney park experience, so diving further into these reviews using data science techniques can provide analytical insight that can have a great benefit to the Company.

Data Explanation

After much searching, I found a dataset on Kaggle that contains over 42,000 TripAdvisor reviews from visitors of the Disney parks in California, Paris, and Hong Kong. The elements of the dataset are as follows (Chillar, 2020):

1. Review_ID: the unique ID given to each review for anonymity and data privacy
2. Rating: the reviewer's rating of their Disneyland experience ranging from 1 (unsatisfied) to 5 (satisfied)
3. Year_Month: the month and year when the reviewer visited the theme park
4. Reviewer_Location: the country of origin of the visitor
5. Review_Text: the review/comment made by the visitor regarding their Disneyland trip
6. Disneyland_Branch: the location of the Disneyland park the reviewer visited

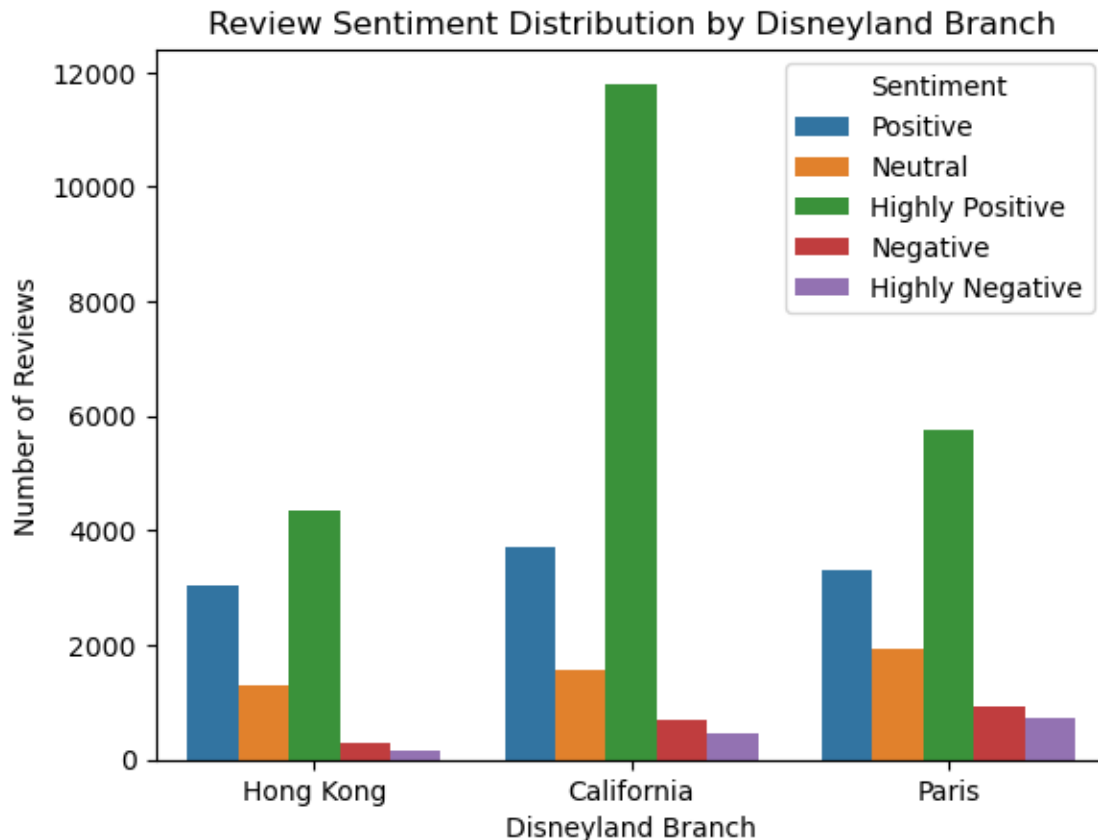
Looking at the dataset variables, there are a few things that can be done to achieve the intended results while also allowing many insights to be seen from the data by way of graphical visualizations and model performance metrics that will ultimately benefit the Walt Disney Company and guest satisfaction. The key variable is review sentiment, which will be derived from the Rating variable present in the dataset.

Questions that I believe end users would be interested in knowing about the guests' sentiment concerning Disney parks are vast, but there are a few key ideas that come to mind that the general public may inquire as to how this study will use review data to benefit the park operations:

1. Does this study consider reviews for every Disney resort in the world today?
2. What do the reviews cite as the most positive part about visiting a Disney park?
3. What do reviewers share about Disney parks that frustrate them the most, prompting negative reviews?
4. How is sentiment determined from reviewer ratings 1 through 5?
5. Guests from which countries leave the most reviews on Disney park experiences?
6. Which Disney park has the highest average monthly satisfaction rating?
7. What are the predictive models crafted in this study supposed to accomplish?
8. Why are techniques such as removing stopwords and porter stemming used?
9. Why isn't the accuracy metric the only performance value used in this research?
10. What can the Company do with this study's results should they be viable?

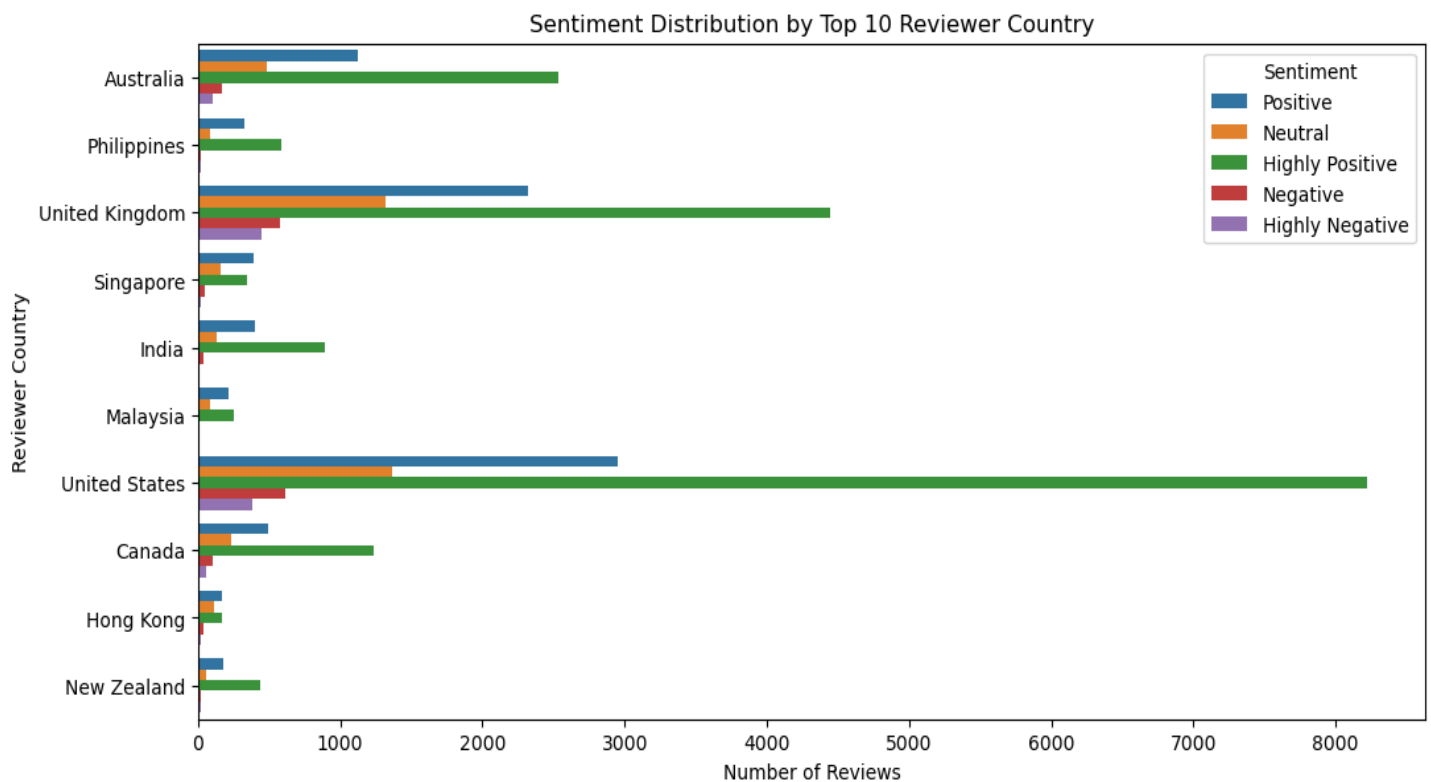
Methods

Many transformations were performed on the data to be ready for the initial exploratory data analysis phase and the model creation stage. After handling missing values, asserting proper data types, dropping irrelevant variables, splitting columns into separate and more useful variables, creating new features, and renaming variables for better readability, some basic insights could be realized. I wished to see the distribution of reviews between each of the three parks as well as the distribution of review sentiment classes. Since the only sign of sentiment in the dataset was the Rating variable, I converted the five rating classes (1 through 5) into categorical string classes (highly negative, negative, neutral, positive, and highly positive). Upon creating the below bar chart, I was able to see the difference in the pattern of reviews between each Disneyland branch within the dataset:



There are clearly many more highly positive and positive reviews within the dataset than the other sentiment classes (this speaks to the Company's efforts to provide a magical experience to the guest). Unfortunately, this class imbalance can be seen as a potential issue that may affect the predictive performance of the forthcoming machine learning algorithms, so hopefully that can be addressed.

With one of the variables featuring geographic location (reviewer's country of origin), I created a chart looking at those countries most represented by their citizens' reviews. The top ten most represented countries are shown as many countries were represented (this speaks to the Company's culture of diversity and inclusion).



It can be seen that reviewers from the United States dominate the data, with the United Kingdom and Australia taking second and third, respectively. Adding in the sentiment class distribution among the countries also allows the identification of cultural opinions.

Once the exploratory data analysis was complete, a preliminary predictive analysis was performed using TextBlob to attach new sentiment class values based on the current review text prior to any textual preprocessing techniques. At this point it was decided to split the parent dataset into three separate datasets (one for each Disneyland branch) to begin addressing some of the class imbalance between highly positive and positive reviews against the other three classes. Sentiment polarity was the score that determined the review's new sentiment classification, with the five string class monikers staying the same and being defined by a range of polarity. The accuracy metric between the actual listed sentiment class values and the TextBlob predicted values was called in addition to a classification report showcasing the precision, recall, and F-1 scores. While it was interesting to see how a quick text sentiment algorithm looked over the data to output three sets of performance metrics, TextBlob served as a baseline model as every metric was abysmal, with twenty-eight percent being the highest accuracy statistic among the three branch models (this statistic belonging to Hong Kong's TextBlob model).

To dive into more complex predictive models with the intent of performing better than the TextBlob sentiment analysis models, textual preprocessing needed to happen here. For this research, the lowercasing of the review text and the removal of all special characters and punctuation was imperative to make sure that only the words themselves were considered in the review's sentiment. Afterwards, the removal of less impactful words known as stopwords took place. These stopwords, such as "the," "of," "a," "and," and others are not seen as significant and/or relevant to the sentiment of the review text, so removing them keeps the essence of the sentiment with little noise to affect the machine learning algorithm's training

process. Stemming the words via the PorterStemming() function finalizes the textual preprocessing sequence by deleting any unnecessary prefixes or suffixes, leaving only the core of the most impactful words to the review's sentiment.

Moving to the training and test set splitting portion of the study, yet another textual analysis technique was to be considered: Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. After choosing the variables within the three datasets to be the training and test set variables (the actual sentiment class values originating from the Rating variable were chosen as opposed to the TextBlob new sentiment values), I split the three Disneyland branch datasets into their respective training and test sets. It was at this point that TF-IDF vectorization was performed to turn the reviews into numerical values indicative of an aggregate of a word's importance relative to the entire review. With the three Disneyland branch datasets split, fit, and transformed for sentiment analysis machine learning models to train on them, it has been decided that a logistic regression, Naïve Bayes, random forest classifier, and an extreme gradient boosting classifier model all need to be generated for each of the three sets.

Model Creation and Analysis

The logistic regression model lends itself to a one-versus-rest classification approach, which looks to be beneficial for the data. The class imbalance was once again taken into consideration with this model as logistic regression has a parameter that balances the weight of each class within the data for more well-rounded accuracy. Unfortunately, the model's results did not suggest such well-roundedness as seen below:

Disneyland California Logistic Regression Model Accuracy: 0.58775061796209
84

Disneyland California Logistic Regression Model Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.28	0.44	0.34	87
Highly Positive	0.85	0.70	0.77	2372
Negative	0.21	0.29	0.25	136
Neutral	0.28	0.37	0.32	326
Positive	0.30	0.39	0.34	720
accuracy			0.59	3641
macro avg	0.38	0.44	0.40	3641
weighted avg	0.65	0.59	0.61	3641

Disneyland Paris Logistic Regression Model Accuracy: 0.562820007877117

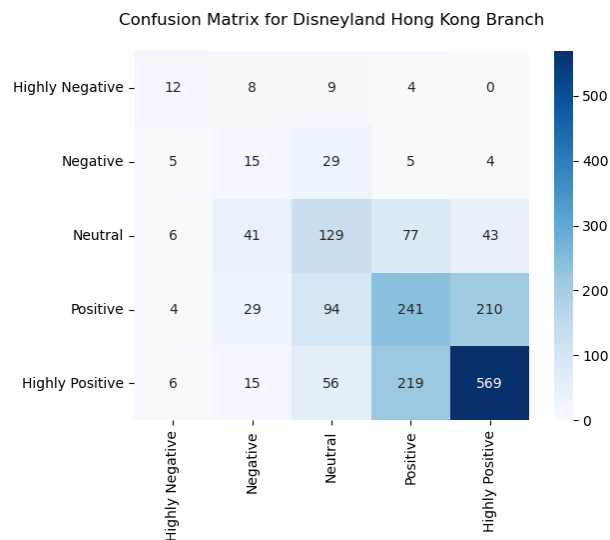
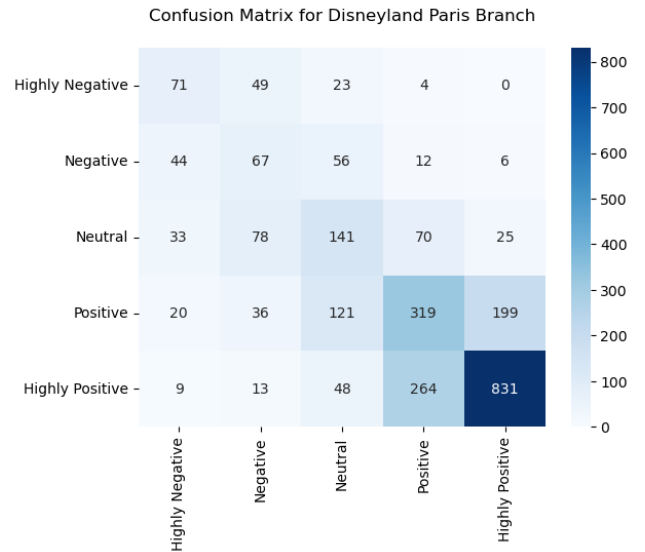
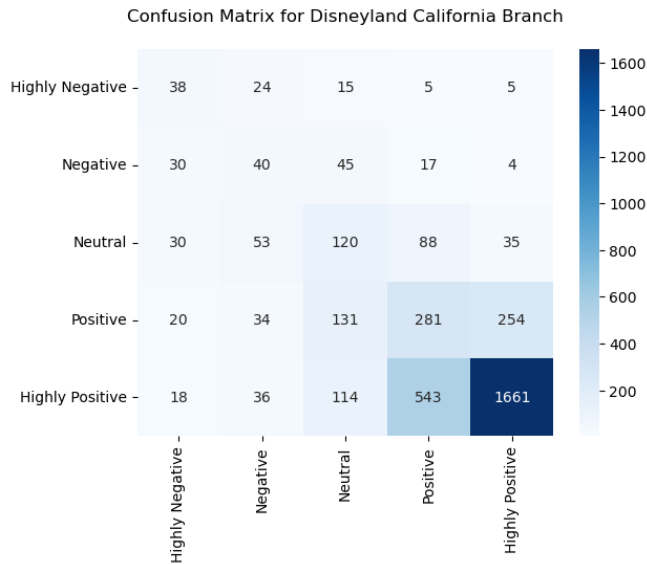
Disneyland Paris Logistic Regression Model Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.40	0.48	0.44	147
Highly Positive	0.78	0.71	0.75	1165
Negative	0.28	0.36	0.31	185
Neutral	0.36	0.41	0.38	347
Positive	0.48	0.46	0.47	695
accuracy			0.56	2539
macro avg	0.46	0.48	0.47	2539
weighted avg	0.58	0.56	0.57	2539

Disneyland Hong Kong Logistic Regression Model Accuracy: 0.527868852459016
4

Disneyland Hong Kong Logistic Regression Model Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.36	0.36	0.36	33
Highly Positive	0.69	0.66	0.67	865
Negative	0.14	0.26	0.18	58
Neutral	0.41	0.44	0.42	296
Positive	0.44	0.42	0.43	578
accuracy			0.53	1830
macro avg	0.41	0.43	0.41	1830
weighted avg	0.54	0.53	0.53	1830



As seen by the results of the logistic models for each Disneyland branch, both the accuracy metric and the confusion matrix show improvement to the TextBlob model analysis.

I'd like to point out the accuracy metrics of the TextBlob sentiment analyzer:

Disneyland California: 18.92% accuracy

Disneyland Paris: 24.53% accuracy

Disneyland Hong Kong: 28.29% accuracy

With the logistic regression model's creation for each Disneyland branch with a balanced class weight factored in, the accuracy has improved for each branch as shown below:

Disneyland California: 58.78% accuracy

Disneyland Paris: 56.28% accuracy

Disneyland Hong Kong: 52.79% accuracy

While these improvements are significant along with the precision, recall, and F-1 score statistics showing similar increases, these models are simply not ready for deployment, so I will move along to the next model on the list: multinomial Naive Bayes.

With the creation of the Naïve Bayes model, the inclusion of the grid search cross-validation method using a parameter grid and search space made for a more optimal model. I will show the results here (I will refrain from showing any more confusion matrices as they all look almost identical from here on out).

Best parameters for the Disneyland California model: {'alpha': 0.1, 'fit_prior': True}

Disneyland California Naive Bayes Model Accuracy: 0.6553144740455918

Disneyland California Naive Bayes Model Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.00	0.00	0.00	87
Highly Positive	0.67	0.99	0.80	2372
Negative	0.00	0.00	0.00	136
Neutral	0.33	0.02	0.03	326
Positive	0.19	0.03	0.06	720
accuracy			0.66	3641
macro avg	0.24	0.21	0.18	3641
weighted avg	0.51	0.66	0.54	3641

Best parameters for the Disneyland Paris model: {'alpha': 0.1, 'fit_prior': False}

Disneyland Paris Naive Bayes Model Accuracy: 0.5395825128003151

Disneyland Paris Naive Bayes Model Classification Report:					
	precision	recall	f1-score	support	
Highly Negative	0.44	0.07	0.13	147	
Highly Positive	0.69	0.78	0.73	1165	
Negative	0.23	0.04	0.07	185	
Neutral	0.33	0.42	0.37	347	
Positive	0.41	0.42	0.42	695	
accuracy			0.54	2539	
macro avg	0.42	0.35	0.34	2539	
weighted avg	0.52	0.54	0.51	2539	

Best parameters for the Disneyland Hong Kong model: {'alpha': 0.1, 'fit_prior': True}

Disneyland Hong Kong Naive Bayes Model Accuracy: 0.5136612021857924

Disneyland Hong Kong Naive Bayes Model Classification Report:					
	precision	recall	f1-score	support	
Highly Negative	0.00	0.00	0.00	33	
Highly Positive	0.56	0.86	0.68	865	
Negative	0.00	0.00	0.00	58	
Neutral	0.63	0.06	0.11	296	
Positive	0.37	0.30	0.34	578	
accuracy			0.51	1830	
macro avg	0.31	0.25	0.22	1830	
weighted avg	0.48	0.51	0.44	1830	

With the performance metrics from the Naive Bayes models, we can see the differences between the results from the logistic regression model. The accuracy metrics from the logistic regression model and the Naive Bayes model are shown below:

Logistic Regression Model

Disneyland California: 58.78% accuracy

Disneyland Paris: 56.28% accuracy

Disneyland Hong Kong: 52.79% accuracy

Naive Bayes Model

Disneyland California: 65.53% accuracy

Disneyland Paris: 53.96% accuracy

Disneyland Hong Kong: 51.37% accuracy

It is interesting to note that Disneyland California's model accuracy increased, yet the other two branches had their model's accuracy decrease from logistic regression to Naive Bayes. It seems that the simplicity of these two models is not enough to handle the imbalanced number of highly positive and positive reviews, as the precision and recall metrics are rather high for these review classes, yet are very low for the other classes. The next model I wish to focus on is the random forest model, one that can handle many different regression and classification problems well.

Since the random forest classifier model is very complex in nature and requires much more computational power than my local device has, I started to experience the drawbacks of weak processing power. To combat this, I defined the major parameters needed to craft the random forest classifiers for each Disneyland branch without creating a grid search. The results of the random forest model are highlighted below:

Disneyland California Random Forest Classifier Accuracy: 0.653117275473771

Disneyland California Random Forest Classifier Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.00	0.00	0.00	87
Highly Positive	0.65	1.00	0.79	2372
Negative	0.00	0.00	0.00	136
Neutral	0.00	0.00	0.00	326
Positive	0.40	0.01	0.02	720
accuracy			0.65	3641

macro avg	0.21	0.20	0.16	3641
weighted avg	0.51	0.65	0.52	3641

Disneyland Paris Random Forest Classifier Accuracy: 0.4954706577392674

Disneyland Paris Random Forest Classifier Classification Report:

	precision	recall	f1-score	support
Highly Negative	1.00	0.01	0.03	147
Highly Positive	0.52	0.98	0.68	1165
Negative	0.25	0.01	0.01	185
Neutral	0.35	0.11	0.17	347
Positive	0.33	0.11	0.17	695
accuracy			0.50	2539
macro avg	0.49	0.24	0.21	2539
weighted avg	0.45	0.50	0.38	2539

Disneyland Hong Kong Random Forest Classifier Accuracy: 0.5207650273224044

Disneyland Hong Kong Random Forest Classifier Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.00	0.00	0.00	33
Highly Positive	0.55	0.92	0.68	865
Negative	0.00	0.00	0.00	58
Neutral	0.46	0.04	0.07	296
Positive	0.42	0.25	0.32	578
accuracy			0.52	1830
macro avg	0.29	0.24	0.21	1830
weighted avg	0.47	0.52	0.44	1830

With the performance metrics from the random forest classifier models, we can see the differences between the results from the logistic regression models and Naive Bayes models.

The accuracy metrics from the logistic regression models, the Naive Bayes models, and the random forest classifier models are shown below:

Logistic Regression Model

Disneyland California: 58.78% accuracy

Disneyland Paris: 56.28% accuracy

Disneyland Hong Kong: 52.79% accuracy

Naive Bayes Model

Disneyland California: 65.53% accuracy

Disneyland Paris: 53.96% accuracy

Disneyland Hong Kong: 51.37% accuracy

Random Forest Classifier Model

Disneyland California: 65.31% accuracy

Disneyland Paris: 49.55% accuracy

Disneyland Hong Kong: 52.08% accuracy

Looking at the three models currently having their results displayed, the random forest classifier model does not seem to offer better performance as a whole than the Naive Bayes model. The only random forest classifier model that performed better than the previous one is the Disneyland Hong Kong model, but not by any significant percentage. Upon seeing the results of these first three predictive models, the Naive Bayes model still outputs the highest average accuracy, yet so far we have seen nothing worthy of deployment. Combined with the poor accuracy statistics, each of the precision, recall, and F-1 scores for every sentiment class other than Highly Positive are very low, with some of the classes receiving zeroes for their subsequent metrics. As it stands now, the only hope for attaining a potentially viable predictive model is the gradient boosting classifier model.

To speed up the computation of the final models and include both Lasso and Ridge regularization, the extreme gradient boosting classifier algorithm is the better choice as opposed to regular gradient boosting classifiers. This gradient boosting algorithm boasts better performance on larger datasets, allows for more parameters for tuning, and lets computation happen in a parallel nature. Here's what the results show:

Disneyland California XGBClassifier Model Accuracy: 0.660258170832189

Disneyland California XGBClassifier Model Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.32	0.07	0.11	87
Highly Positive	0.68	0.99	0.81	2372
Negative	0.21	0.02	0.04	136
Neutral	0.31	0.03	0.06	326
Positive	0.28	0.05	0.08	720
accuracy			0.66	3641
macro avg	0.36	0.23	0.22	3641
weighted avg	0.54	0.66	0.55	3641

Disneyland Paris XGBClassifier Model Accuracy: 0.5340685309176841

Disneyland Paris XGBClassifier Model Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.47	0.18	0.26	147
Highly Positive	0.58	0.92	0.71	1165
Negative	0.38	0.06	0.10	185
Neutral	0.45	0.23	0.30	347
Positive	0.40	0.23	0.30	695
accuracy			0.53	2539
macro avg	0.45	0.33	0.33	2539
weighted avg	0.49	0.53	0.47	2539

Disneyland Hong Kong XGBClassifier Model Accuracy: 0.514207650273224

Disneyland Hong Kong XGBClassifier Model Classification Report:

	precision	recall	f1-score	support
Highly Negative	0.67	0.06	0.11	33
Highly Positive	0.55	0.88	0.68	865
Negative	1.00	0.02	0.03	58
Neutral	0.41	0.08	0.14	296
Positive	0.40	0.26	0.32	578
accuracy			0.51	1830

macro avg	0.61	0.26	0.25	1830
weighted avg	0.50	0.51	0.44	1830

With the performance metrics from the extreme gradient boosting models, we can see the differences between the results from the three previous models. The accuracy metrics from all four classifier models are shown below:

Logistic Regression Model

Disneyland California: 58.78% accuracy

Disneyland Paris: 56.28% accuracy

Disneyland Hong Kong: 52.79% accuracy

Naive Bayes Model

Disneyland California: 65.53% accuracy

Disneyland Paris: 53.96% accuracy

Disneyland Hong Kong: 51.37% accuracy

Random Forest Classifier Model

Disneyland California: 65.31% accuracy

Disneyland Paris: 49.55% accuracy

Disneyland Hong Kong: 52.08% accuracy

Extreme Gradient Boosting Classifier Model

Disneyland California: 66.03% accuracy

Disneyland Paris: 53.41% accuracy

Disneyland Hong Kong: 51.42% accuracy

As it stands, the Naive Bayes model and the extreme gradient boosting classifier model accuracy metrics are the best of the four models. Out of these two, the extreme gradient boosting model exhibits the better-performing classification report, so we can declare this predictive model the best of the bunch.

Recommendations/Implementation Plan

Now that all the models have been crafted, it is clear that none of these predictive models are accurate enough to be considered for deployment. Each classification report printed has been lackluster in accurately predicting imbalanced class review sentiment. There are many ways to address this in the future to increase accuracy, however, without more computational power and memory, it may take too long to process. Such techniques to potentially increase accuracy, precision, recall, and F-1 scores are as seen below:

1. Run the previous predictive models using TextBlob's new review text sentiment values
2. Enforce more hyperparameter tuning techniques specific to the class imbalance
3. Incorporate class balancing methods by either duplicating the minority class values to equal the number of majority data points or dropping enough majority values to equal the number of the minority class values

While the benefits of sentiment analysis to businesses are many, including the improvement of brand sentiment, customer service, and competitor analysis, it looks rather difficult to explore

these benefits with the models and the data provided in this study (Blogger, 2023). Even more ambiguous would it be to construct an implementation plan of such sentiment analysis in production as none of the four models across Disneyland California, Paris, and Hong Kong reviews have yielded a potentially deployable solution. It is highly recommended that if this data were to be analyzed with any sort of accurate predictive power, the sentiment class imbalance would need to be properly addressed while increasing computational processing abilities.

Assumptions

One of the biggest assumptions made within this research is that the thresholds for sentiment polarity via TextBlob should have been equal for each sentiment class when the number of reviews in each class was heavily skewed in favor of Highly Positive and Positive reviews. The issue of class imbalance could have been scrutinized further and would have potentially ended up being solved provided the TextBlob new sentiment classes were much more accurate than they were within the realm of this study. It may have even proven more effective to use the new sentiment class observations from TextBlob instead of sticking with the original classes.

Challenges/Limitations

Venturing into complex studies such as this allows one to potentially discover a solution to a business problem that can benefit many guests, end users, developers, and the business as a whole. However, this project has shown me the limitations of my current local device and the power of big data and fast, powerful computational processing ability. For both the random

forest classifier and the extreme gradient boosting classifier models, I had to initialize each model with set parameters just to output a model at all. I waited over an hour to output the three Disneyland branch random forest classifier models using a grid search cross-validation method, and after nothing had been generated, it was clear that I would have to scale down the models to receive results. While it was a struggle to find out that my research was beyond my local device's computational power, it was gratifying to see that entering the world of data science from a business perspective allows producing projects of a much higher quality at faster speeds.

Ethical Assessment

Ethics are always a concern when dealing with data; this research is no different. An apparent ethical concern is that I currently work for the Company, thereby asserting some form of bias to be recognized and set aside to be an impartial viewer of the data to report results that are not obscured by my direct affiliation with the Company. Adding to this is the privacy of the users involved in the reviews featured within the dataset. While their identity is to be protected and/or anonymized for data protection purposes, it seems that only some measures have been taken, and more could be done to prioritize the privacy of the review creators. While the variable outlining the reviewer's location could provide definitive insight as to the sentiment of reviewers in a specific region of the world, the reviewers could potentially be reidentified by combining the location, the review contents, and the year/month it was published.

Another data ethics discussion relating to this study is the fact that these reviews came from TripAdvisor, which brings into question the website's data usage policies and whether the reviewers had been notified that their reviews could be used for educational research like this. TripAdvisor is one of the most trusted hospitality/tourism websites out there, and for them to not tell their users that any reviews posted could be used for market research would be a big concern. Checking TripAdvisor's policies as they relate to the scraping and usage of their data for research purposes would mitigate that concern and allow the study to progress towards a level mimicking real-world application.

Future Uses/Additional Applications

This research is extremely beneficial to the Company (provided a viable model with satisfactory results is created) as the Company can hone in on the specific elements of the park experience that are favorable and enjoyable to guests as well as those that need extra attention. The word cloud charts made in the EDA phase are rather important because they focus on the negative reviews that mention wait times and waiting as a major issue that needs to be looked into. The bar chart visualizing the top ten countries whose citizens left reviews for the three Disneyland branches shows the countries that care enough about Disney to speak their minds about park best practices and questionable moves. This knowledge could be used in the future to identify these countries' spending habits at the parks and eventually their citizens as high-priority guests, leading to decisions geared towards catering to those countries.

Conclusion

After careful analysis, feature engineering, and training of the data using viable model choices, it has been found that none of the predictive models in this study yielded any useful results based on the performance metrics produced. The Walt Disney Company is a stable company with a bright future ahead according to the data, and with predictive modeling techniques to aid the Company and its current and potential guests, that future should remain bright barring any major park changes or events that balance the review sentiments in a negative way.

Appendix/References

1. Blogger, G. (2023, September 26). Top 5 Benefits of Sentiment Analysis for Businesses. Retrieved November 7, 2024, from <https://determ.com/blog/top-5-benefits-of-sentiment-analysis-for-businesses/>
2. Chillar, A. (2020, January 1). Disneyland Reviews. Retrieved November 7, 2024, from <https://www.kaggle.com/datasets/arushchillar/disneyland-reviews>