Metacritic Game Review Sentiment Analysis

Springboard Data Science Capstone 2 Project
May 26th 2020 Cohort
Filiberto Aguilar

The Problem

- What did gamers enjoy and dislike the most in games on the Xbox One, Playstation 4 and Nintendo Switch consoles?

- What makes a good RPG game appealing to a gamer?

Potential Clients

The Data

- Over 20,000 game reviews were scraped off Metacritic

- 15 reviews were scraped for all game titles across the Xbox One, PS4 and Switch for games with at least 15 reviews

- Main features scraped:
  - Game title
  - Platform
  - Developer
  - Game genre
  - Number of players
  - ESRB rating
  - Release date
  - Review

Data Cleaning

- 2% percent of the data was dropped to remove missing values for various features

- 16% of the data did not have a value for the number of players feature
    - Titles with a missing value were googled and assigned a value of ‘singleplayer’ or ‘multiplayer’
    - All other observations for this feature had their values reduced to ‘singleplayer’ or ‘multiplayer’

- Text data was processed in the following order:
    1. Transformed into lower case
    2. Stripped of digits
    3. Expanded contractions
    4. Emojis transformed into words
    5. Stripped of punctuation
    6. Stripped of white space
    7. Filtered from stop words
    8. Lemmatized

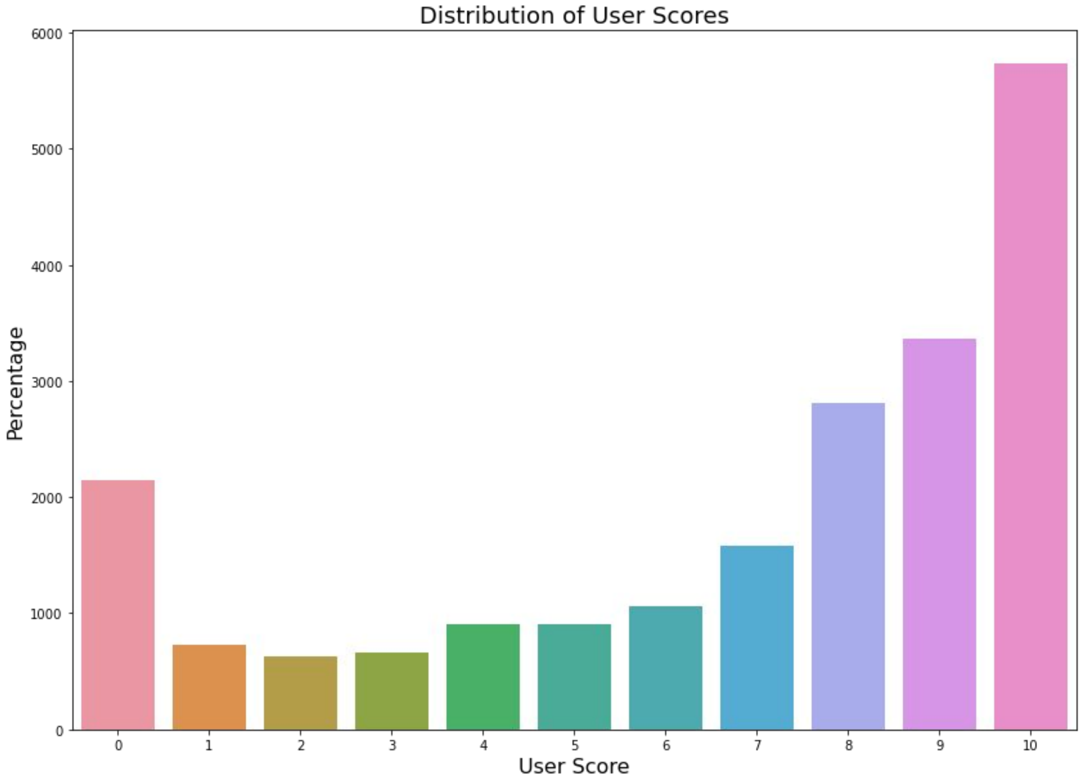
Exploratory Data Analysis

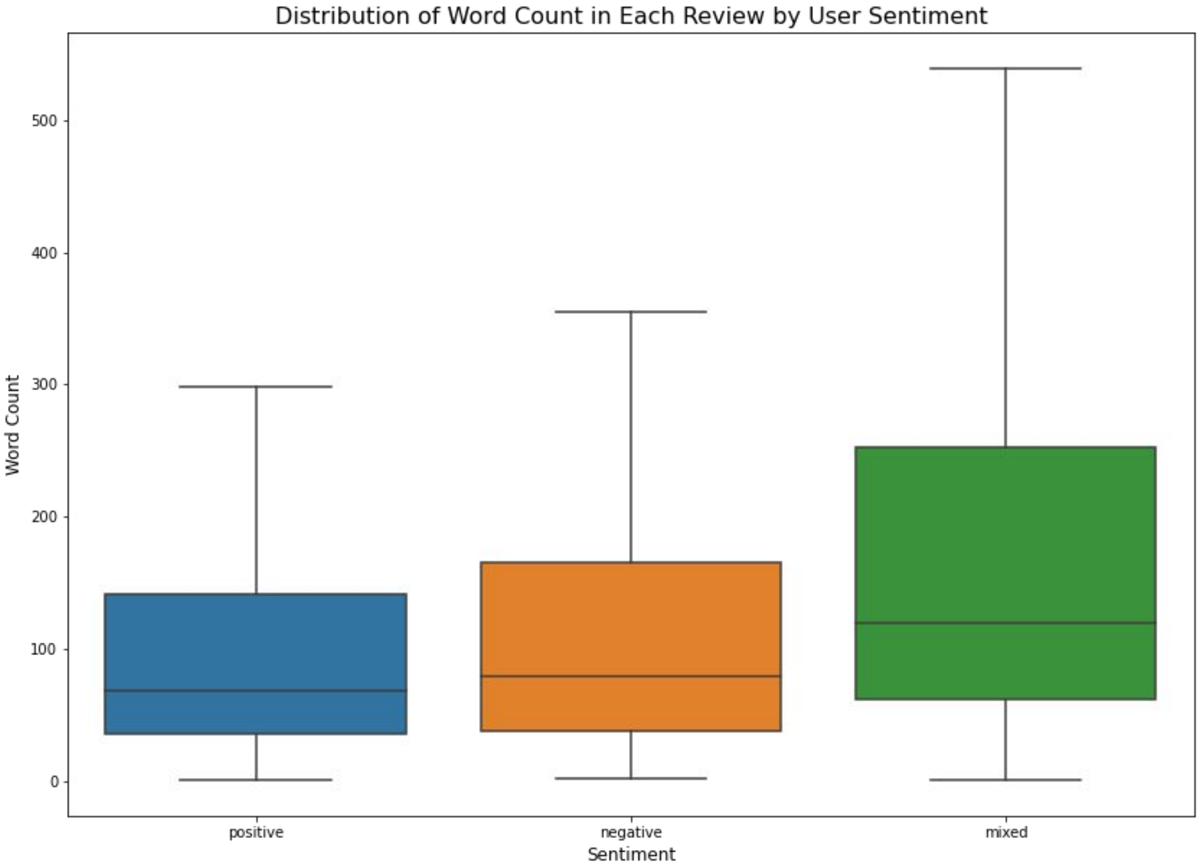
Distribution of User Scores

- The following user score ranges distinguish the user's sentiment:
  - Positive: 8 -10
  - Mixed: 5 - 7
  - Negative: 0 - 4

- Most of the reviews in the dataset were positive accounting for about 57%

- About 10% of all reviews received a 0



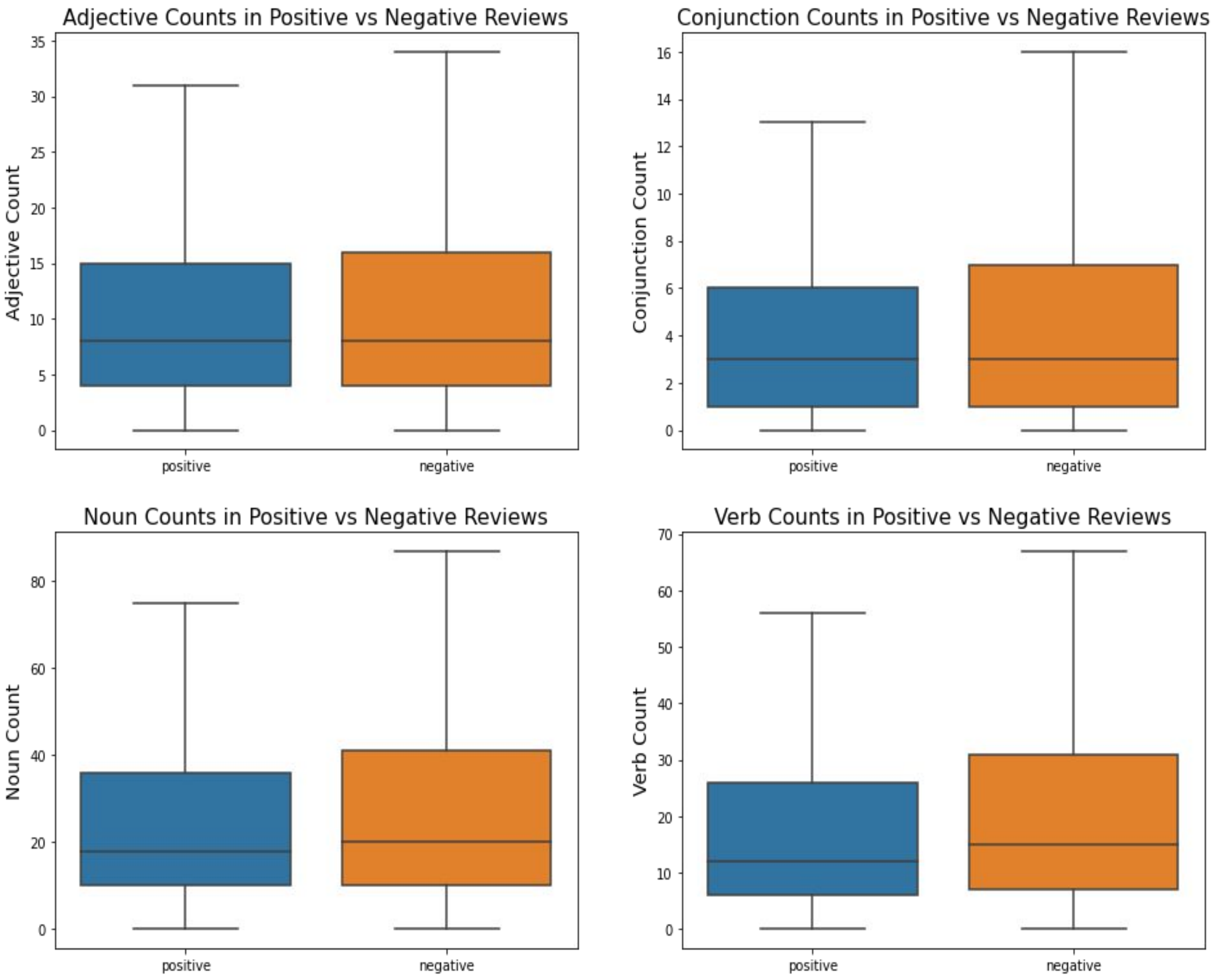


Are positive user reviews longer than negative ones?

- Mixed reviews on average contained the longest reviews
- There was a statistically significant difference among positive and negative reviews (p-value < 0.001)

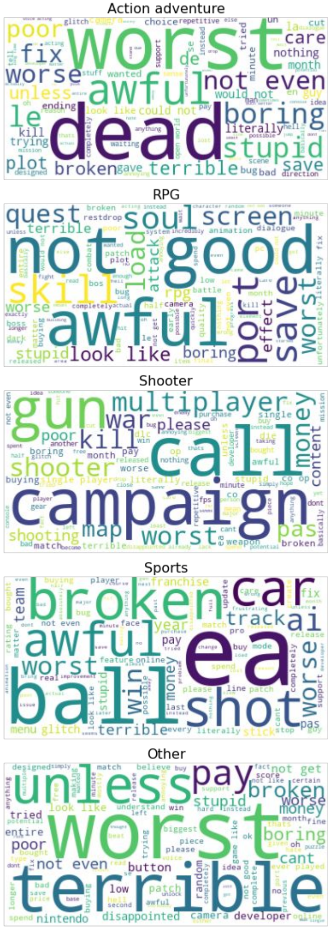
Which is the most common part of speech among positive and negative reviews?

- Nouns on average were the most common POS among reviews
- There was evidence to suggest that positive reviews do not contain more nouns and verbs than negative ones (p-value < 0.001)



Most Predictive Words in Positive Reviews by Genre

Most Predictive Words in Negative Reviews by Genre



What are the most predictive words in game reviews by genre and sentiment?

- Positive action adventure games and RPG were strongly represented by words that seemed to equate a great action adventure game to a cinematic-like experience

- EA is very predictive of negative sport game reviews

Modeling

Six models were considered:

- K-nearest Neighbors (KNN)
- Logistic Regression
- Linear Support Vector Machines (SVM)
- Multinomial Naive Bayes
- Random Forest
- Gradient Boosting

Modeling steps

Trained all models for best minimum document frequency via grid search cross validation:

- Created pipeline to vectorize text; CountVectorizer then TFIDF transformer
- 5 fold cv
- Each model performance was evaluated by the 'ROC-AUC' score
- Selected best two performing models to hypertune

Hypertuned linear SVM and Logistic Regression for inverse of regularization strength:

- 5 fold cv
- Each model performance was evaluated by the 'ROC-AUC' score

Model Performance

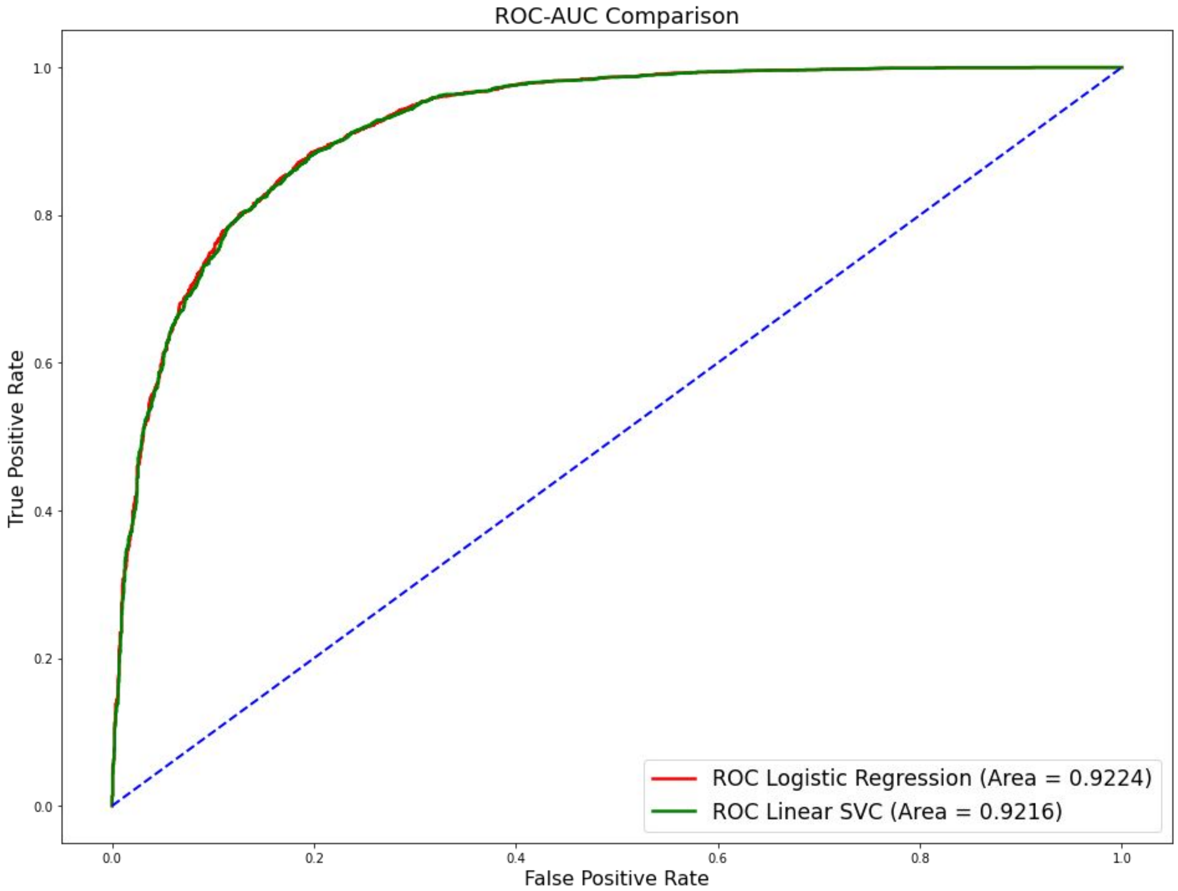
Best parameter inverse of regularization strength value:
Logistic Regression = **1**, **Linear SVM = 0.01**

Model	**ROC-AUC**
KNN	0.521
Logistic Regression	0.923
SVM	0.919
Multinomial Naive Bayes	0.900
Random Forest	0.899
Gradient Boosting	0.866

Best Performers

- From the ROC-AUC curves it is apparent that Logistic Regression model slightly outperformed the Linear SVM

- The decision was made to move forward with the Logistic Regression



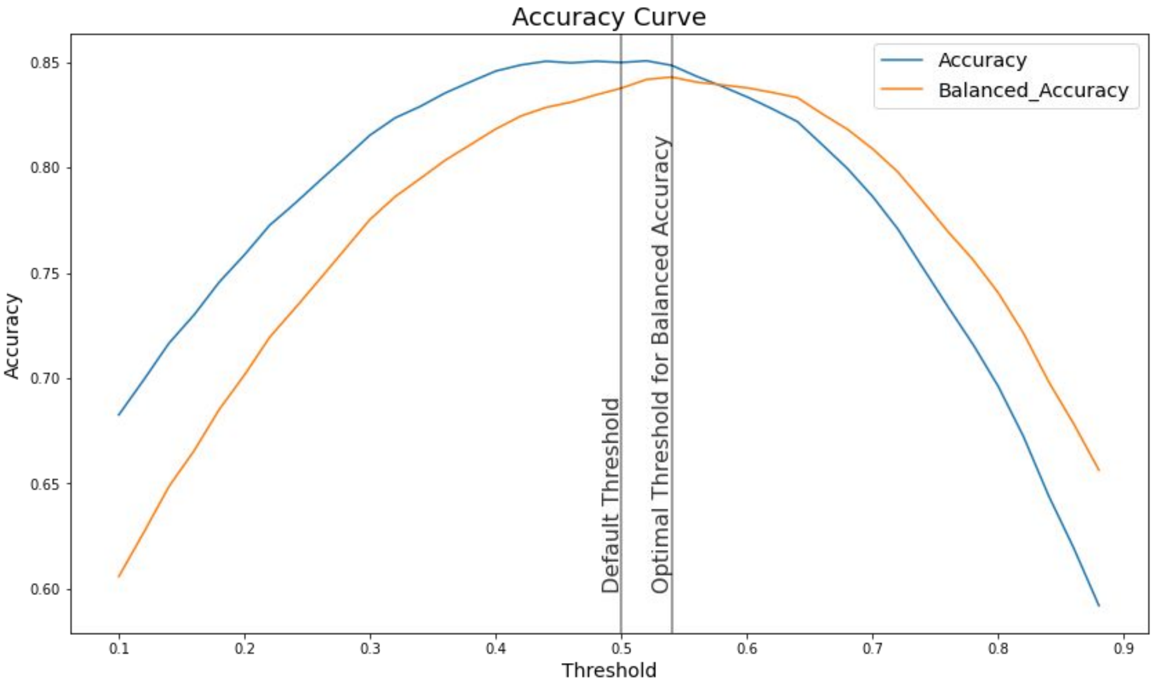
Model Evaluation

Business Case 1

- Potentially useful for companies who want to assess a general response to a game release by customers on social media

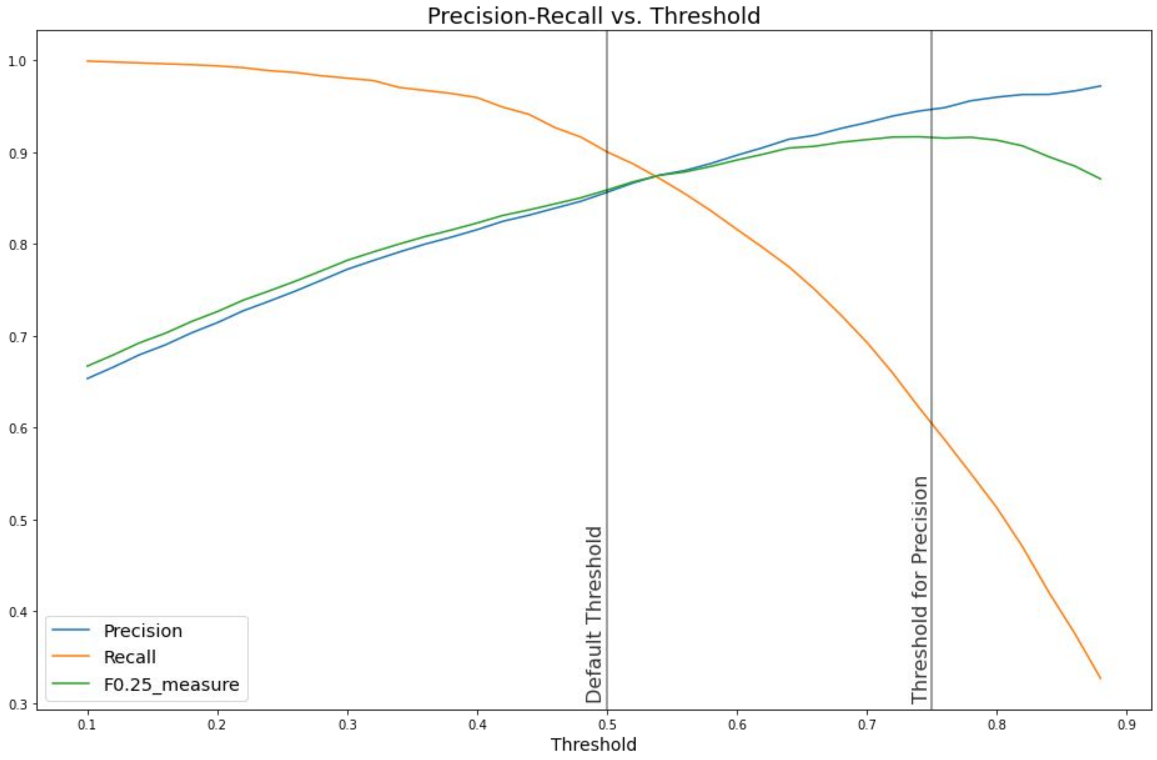
- Unbalanced dataset => model optimized for balanced accuracy

- Optimal threshold value was approximately 0.54



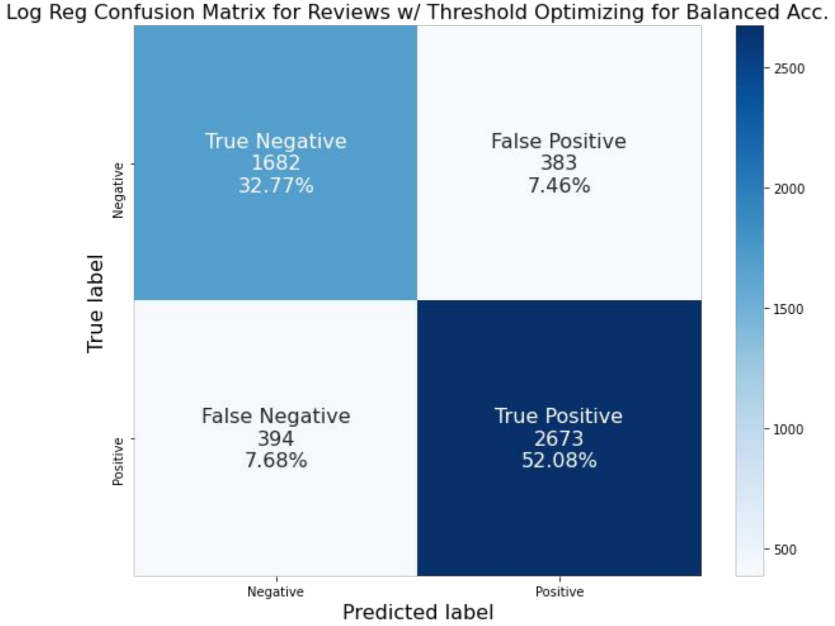
Business Case 2

- Potentially useful for companies who look to find potential influencers on social media that can promote or advertise the product
- Optimized for precision
- F0.25 favored precision metric in thresholding as compared to recall
- Optimal threshold value was approximately 0.75



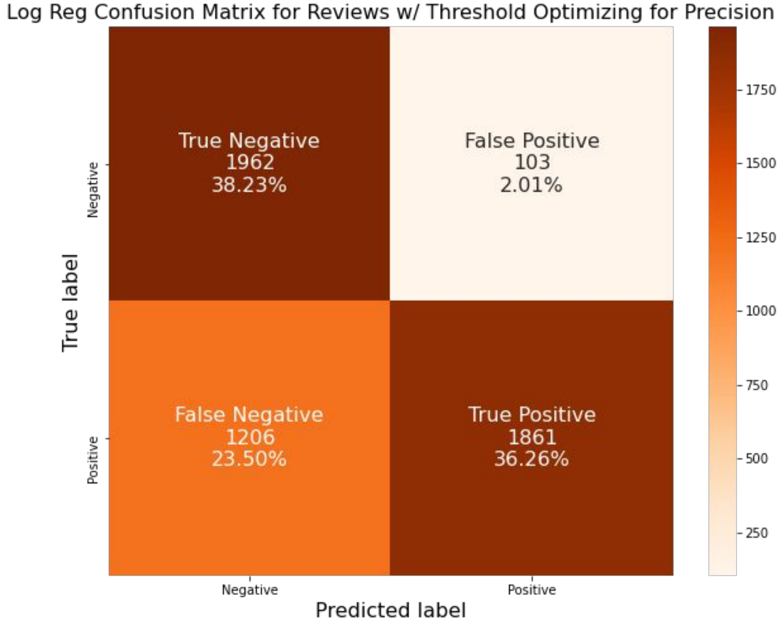
Confusion Matrices

Business Case 1



Accuracy=0.849
Balanced_Accuracy=0.843
Precision=0.875
Recall=0.872
F1 Score=0.873

Business Case 2



Accuracy=0.745
Balanced_Accuracy=0.778
Precision=0.948
Recall=0.607
F_Score=0.917

Conclusion

- Gamers that play RPG games enjoy the combat style and different quests offered in the games, but detest the screen loading and saving time.
- Players are tired or very unsatisfied with sport games developed by EA.
- Best performing model was a Logistic Regression model, achieving an ROC-AUC score of around 92%
  - With a threshold of 0.54 the highest balance accuracy is 84.3%, which allows game developers to accurately assess their games versus their competitors
  - With the adjusted F score and a threshold of 0.75 the model achieved a precision of about 95%, making the model an efficient predictor of positive reviews which can be used to find potential influencers

Special thanks to:

- Benjamin Bell, Springboard mentor
- Springboard community

Sources

Movie Reviews, TV Reviews, Game Reviews, and Music Reviews. (n.d.). Retrieved November 21, 2020, from https://www.metacritic.com/

Sarkar, D. (2018, December 04). A Practitioner's Guide to Natural Language Processing (Part I) - Processing & Understanding Text. Retrieved November 21, 2020, from https://towardsdatascience.com/a-practitioners-guide-to-natural-language-processing-part-i-processing-understanding-text-9f4abfd13e72

Says:, S., Says:, V., Says:, S., Says:, M., Says:, D., Says:, B., . . . Says:, U. (2017, September 13). Basic evaluation measures from the confusion matrix. Retrieved November 21, 2020, from https://classeval.wordpress.com/introduction/basic-evaluation-measures/