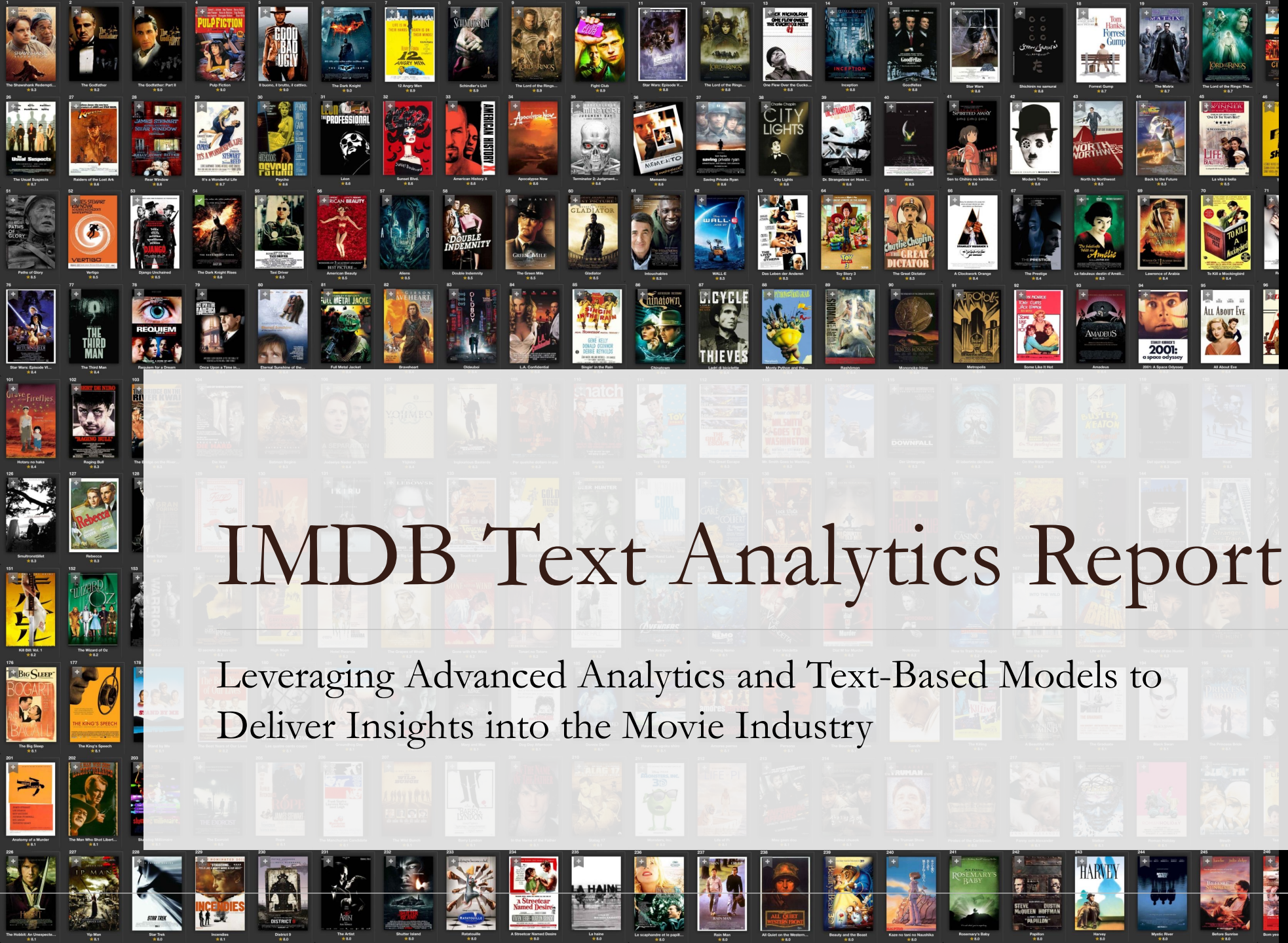
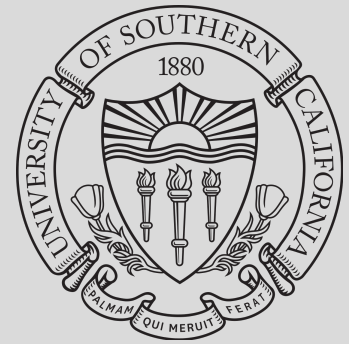


IMDB Text Analytics Report

Leveraging Advanced Analytics and Text-Based Models to Deliver Insights into the Movie Industry



12/17/21

Executive Summary

Internet Movie Database and Datasets
- IMDB currently stores data on over 8 million movies and TV shows with over 80 million registered users
- Parent company is Amazon.com, but operate independently as a private subsidiary

Initial Analysis and Text Preprocessing
- Cleaned text through a regex, stop word removal, and lemmatization pipeline
- Common bi-gram phrases can help differentiate movie genres and find positive/negative phrases

Sentiment Analysis Model
- Sentiment Analysis model had 90% accuracy when predicting positive reviews
- This model was well-balanced, with similar precision and recall scores

Multiclassification Genre Model
- Genre Summary model correctly predicted a single genre with 43% accuracy
- Genre Review Model achieves an ROC-AUC of 0.91 using an MLP classifier

Text-Based Models Offer New Insights Into Box Office Success
- Sentiment Analysis output can deliver automated reports into data products for customer
- Genre Review model can be used as a tool to inform review quality to specific genres
- Genre Overview model can automate tasks and can improve future summaries

Financial Impact

Improved Operational Efficiency

$18 M Return On Investment

Internet Movie Database (IMDB)

One of the Largest Movie Databases

- An online searchable database that stores information on movies, television shows, independent films, production, cast, summaries, reviews, and ratings
- Contains over 8 million titles with 83 million registered users
- IMDB users can vote (1-10) on any title on the database, scores are aggregated to assign a single IMDB rating for that movie
- IMDB ratings are often used as an indicator of a movie's quality and integrated into many different streaming services
- Purchased by Amazon in 1998 for $55 Million, currently operates as a subsidiary, private company
- **Our business proposal was to leverage IMDB's database, summaries, and reviews to automate processes and deliver new insights into audience sentiments and review quality**

Business Scope

As IMDB user-base continues to grow, the volume of user reviews and ratings are going to increase exponentially.

With this large database and detailed scoring system, automating processes through NLP models can improve movie tagging and become a foundation for search recommendation systems.

This can be accomplished through a sentiment model and multiclassification models using the latest deep learning architectures.

These models can also deliver advanced insights into how sentiment and reviews can influence film performance.

Our goal is to deliver four business solutions to address the needs of the company and open new analysis opportunities that can be offered as services to production companies.

Business Solutions

1

Sentiment Analysis Model

This model aims to determine positive/negative sentiment from IMDB reviews. These model scores can be integrated into IMDB's rating system, delivering a more robust representation of viewer sentiment.

2

Multiclassification Genre Review Model

This model aims to predict the genre of the film that a reviewer had seen. By being able to accurately predict the genre, we can determine the "quality" of a review (determining if a review is more general or specific to a movie)

3

Multiclassification Genre Summary Model

This model takes a quick three-five sentence summary and predicts the movie genre. In production, this model can automate our genre tagging process and improve how future summaries are written.

4

Model Outputs as an External Solution

Finally, our three model outputs and scores can be compared with film budget, ratings, and box office performance to identify trends. By exploring these output, we can deliver new solutions to guide production companies into making more informed decisions.

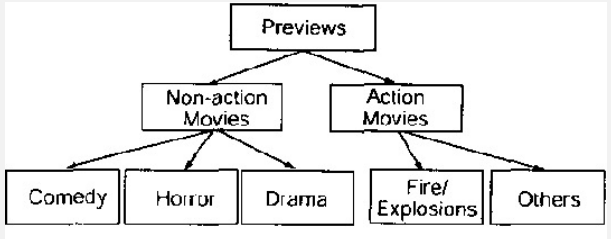
Model Architecture Summary

Sentiment Analysis Model

Use Case: Predicts the review sentiment as positive or negative

Architecture: Recurrent Neural Network with a Long-Term Short Memory (LTSM) layer. Included Embedding layer by training our own Word2Vec model

Output: Gives a score between 0 and 1 with high scores (0.5 or higher) indicating more positive sentiment and low scores (0.5 or lower) indicating negative sentiment



Genre Classification Model

Use Case: Predicts the genre of the movie based on the synopsis

Architecture: Recurrent Neural Network with a Long-Term Short Memory (LTSM) layer. Included Embedding layer by training our own Word2Vec model

Output: Returns a collection of scores that sum to 1. The output with the highest score represents the predicted genre

Review Classification Model

Use Case: Predicts the movie genre of a particular review

Architecture: Multilayer perceptron Neural Network

Output: Returns a collection of scores that sum to 1. The output with the highest score represents the predicted genre. If there is no clear prediction, it is likely these reviews are more "general" in nature.

Data Sources

IMDB Movie Reviews Datasets

IMDB 320,000 Movie Reviews: Contains over 320,000 reviews collected from 1998 to 2021. Includes movie title, genre, ratings, and number of reviews as fields

IMDB Dataset: Contains 50,000 movie reviews with a labeled target column for sentiment analysis (positive or negative)

IMDB Movie Metadata

The Movies Dataset: Contains detailed information on over 45,000 movies. Includes the movie title, genre, budget, release date, box office performance, production companies and a synopsis detailed what the movie is about

Model Text Preprocessing

1 Regex Cleaning

Removed malformed text, irrelevant hanging punctuation

Removed 3+ duplicated characters to get original meaning of word

(Example: helpppp to help)

2 Stop Word Removal

We removed the most common words (the) as they didn't add much new information for our models

Our custom stop words included:

Also, well, much, get, take, make, try, live, come, must, turn, film, movie, story, back

3 Lemmatization

In order to group together inflected forms of words, we used lemmatization. We chose lemmatization over stemming as lemmatization considers word context and keeps a meaningful base form.

Example: studies/studying becoming study

4 N-Grams

For exploratory analysis, we looked at the most popular 2–3-word phrases (n-grams) to draw quick insights into key words that could influence our models' decisions

5 TF-IDF Analysis

Using our n-grams, we determined key short phrases that we associated with sentiment and specific movie genres

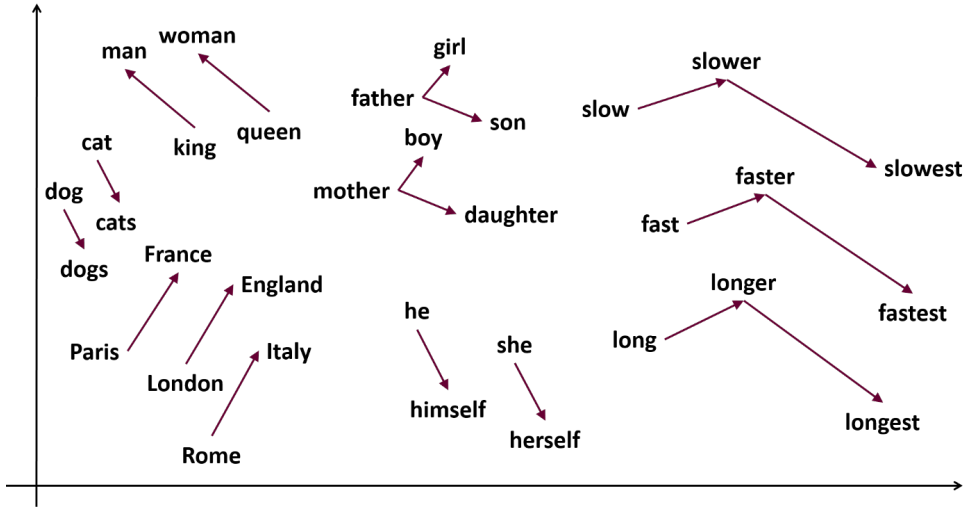
Instead of using a Count Vectorizer, we used TF-IDF calculations to get the relatively important phrases

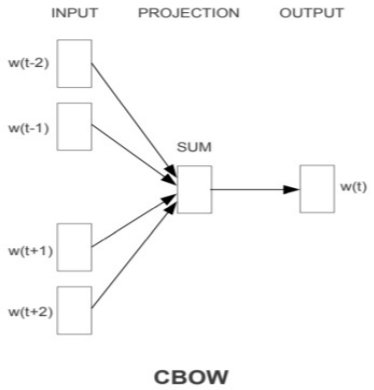
6 Word Embeddings

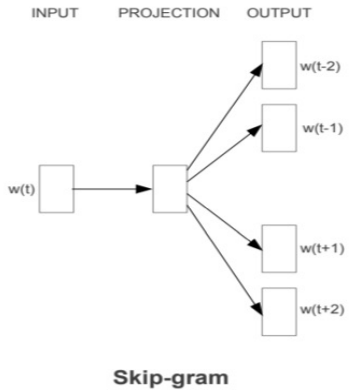
Since we were using deep learning architecture, many of these advanced models require (or can train) and embedding matrix.

Using Word2Vec, we trained some of our own word embeddings for the models

Word Embeddings Explained (Word2Vec)







"You shall know a word by the company it keeps"

– JR Firth

Word embeddings is a technique to transform words into numerical representations by considering characteristics of the words.

Some characteristics might include semantic meaning or context windows with these words transformed into vectors.

The two most common architectures are CBOW (predicts a target word from the surrounding words) and skip-gram (predicts the surrounding words from the target word).

Word2Vec was developed by Google to use these related architectures to create word embeddings that can are highly predictive when used in models.

Source: Implementing Word2Vec in Tensorflow, Word2Vec Part 1 DSO 560 Course Notes, NLP 101: Word2Vec – Skip-gram and CBOW

Exploratory Sentiment Analysis

Positive Sentiment Word Phrase Cloud

The positive sentiment bigrams show some of the common phrases that we typically expect to see for positive sentiment.

Negative Sentiment Word Phrase Cloud

Conversely, looking at negative sentiment we can see common phrases that we would typically use for a film we didn't enjoy.

Exploratory Genre Analysis

This analysis also gives us some intuition into which words may be added to our stop word list

Drama

Average Mean Rating 7.13

Average Number of Reviews 11.29

Some Common Bigram Expressions

Fall Love, Love Story, Good Film, Feel Like, Make Movie, Even Though, Film Not

Thriller/Horror

Average Mean Rating 5.57

Average Number of Reviews 15.9

Some Common Bigram Expressions

Horror Movie, Horror Film, Low Budget, Bad Movie, Take Place, Make Sense, Not Bad

Action /Adventure

Average Mean Rating 6.18

Average Number of Reviews 22.07

Some Common Bigram Expressions

Special Effect, Bad Guy, Action Movie, Good Movie, Martial Art, Fight Scene

Comedy

Average Mean Rating 6.88

Average Number of Reviews 15.54

Some Common Bigram Expressions

Romantic Comedy, High School, Fall Love, Like Movie, Movie Not, Like Movie, Even Though

Movie Review Genre Classifications

Comedy

Average Mean Rating 6.76

Total Number of Reviews: 16606 (25%)

Others

Average Mean Rating 6.46

Total Number of Reviews: 12288 (19%)

Action

Average Mean Rating 6.11

Total Number of Reviews: 14051 (22%)

Crime

Average Mean Rating 6.61

Total Number of Reviews: 5236 (8%)

Drama

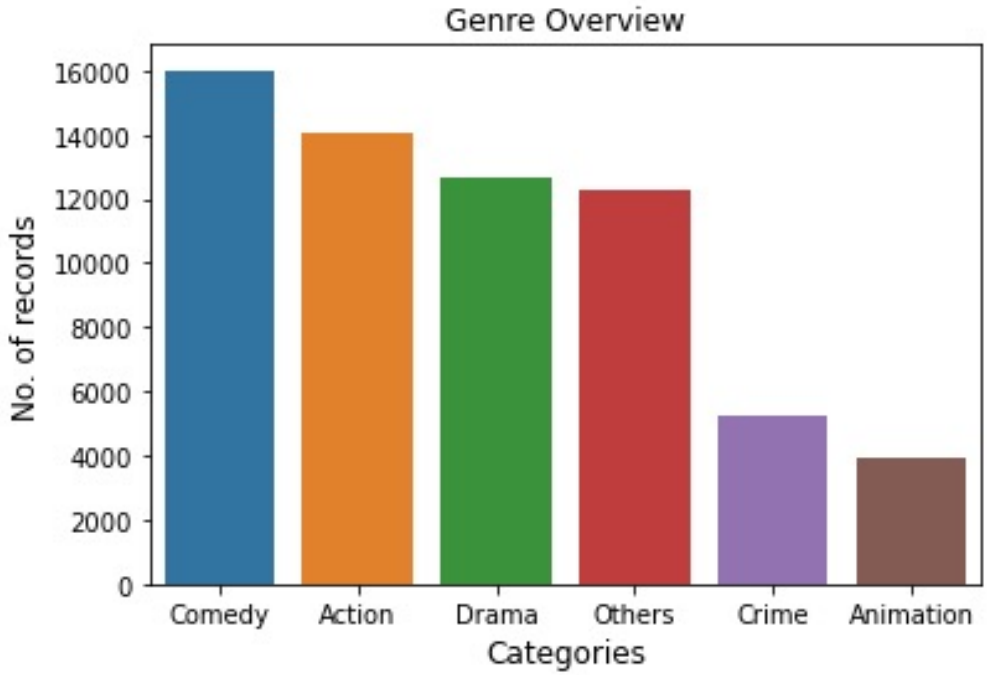
Average Mean Rating 6.87

Total Number of Reviews: 12624 (20%)

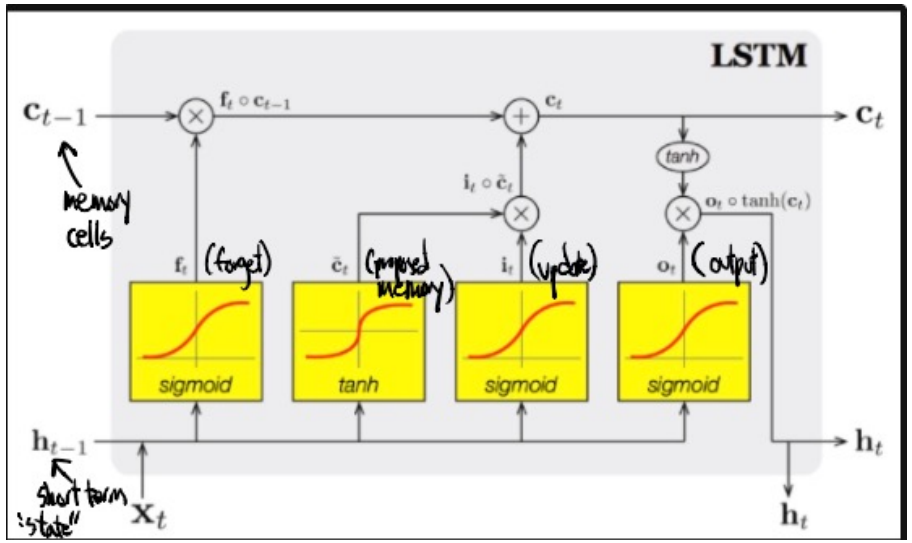
Animation

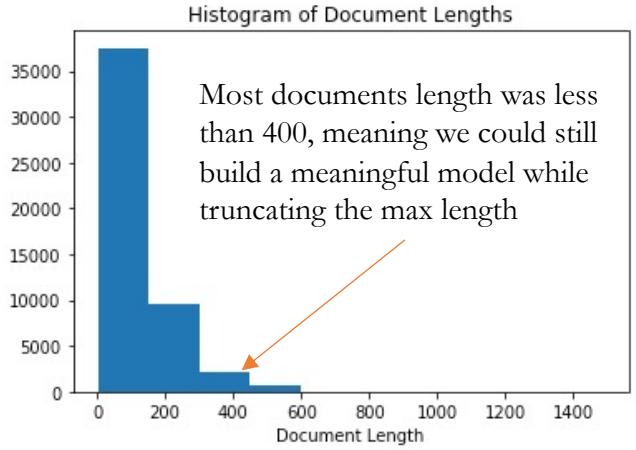
Average Mean Rating 7.00

Total Number of Reviews: 3944 (6%)



Sentiment Analysis Model Architecture





Model Architecture:

This model was composed using our own trained Word2Vec 300 vector embedding with a LSTM layer. Long-term short memory is a special kind of RNN architecture that are designed to remember information for long periods of a time.

The model outputs the probability(using a sigmoid function) that the written review has positive sentiment.

Model Hyperparameters:

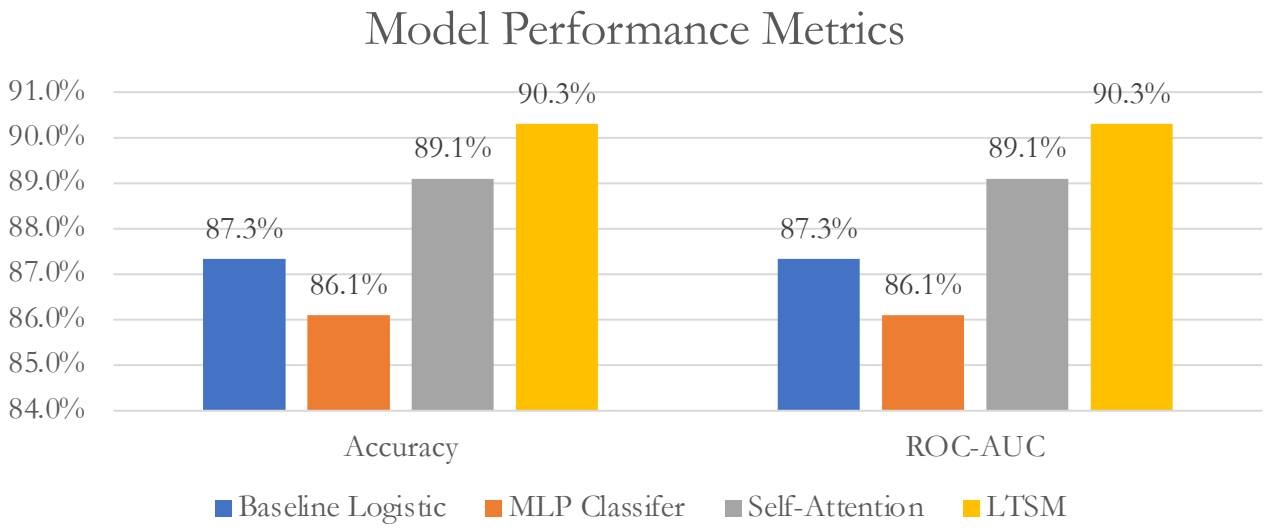
Voacb Size: 10,000, Max Length = 300

Layers: Embedding (Using Trained Word2Vec Model), Masking, LSTM (64 units), Dense (16 Units), Dense (1 Unit Sigmoid)

Units SoftMax)

Sentiment Analysis Model Performance

- Our ML models perform significantly better than random guess, with a **lift of 40.3%**
- The deep learning architectures appear to have the best performance, with a 90% accuracy and 0.90 ROC
- The model has a good balance between precision and recall, indicating it is not bias towards a positive or negative prediction
- Our recommendation would be to use this model to automate reports and present a detailed customer sentiment analysis to production companies packaged as a new data product



Confusion Matrix	Actual Positive	Actual Negative
True Positive	2210	270
True Negative	241	2279

Sentiment Analysis Model Output

Correct Prediction: Positive

Original Classification: Animation, Family, Comedy

"This a **fantastic movie** of three prisoners who become famous. One of the actors is george clooney and I'm not a fan but this roll is **not bad**. **Another good thing** about the movie is the soundtrack (The man of constant sorrow). **I recommand this movie to everybody**. Greetings Bart"

Predictions: Positive 95%

Although this review uses phrases like I'm not a fan and bad, the model is able to correctly pick out that in this context, they weren't suggesting a negative tone towards the film..

Incorrect Prediction: Negative

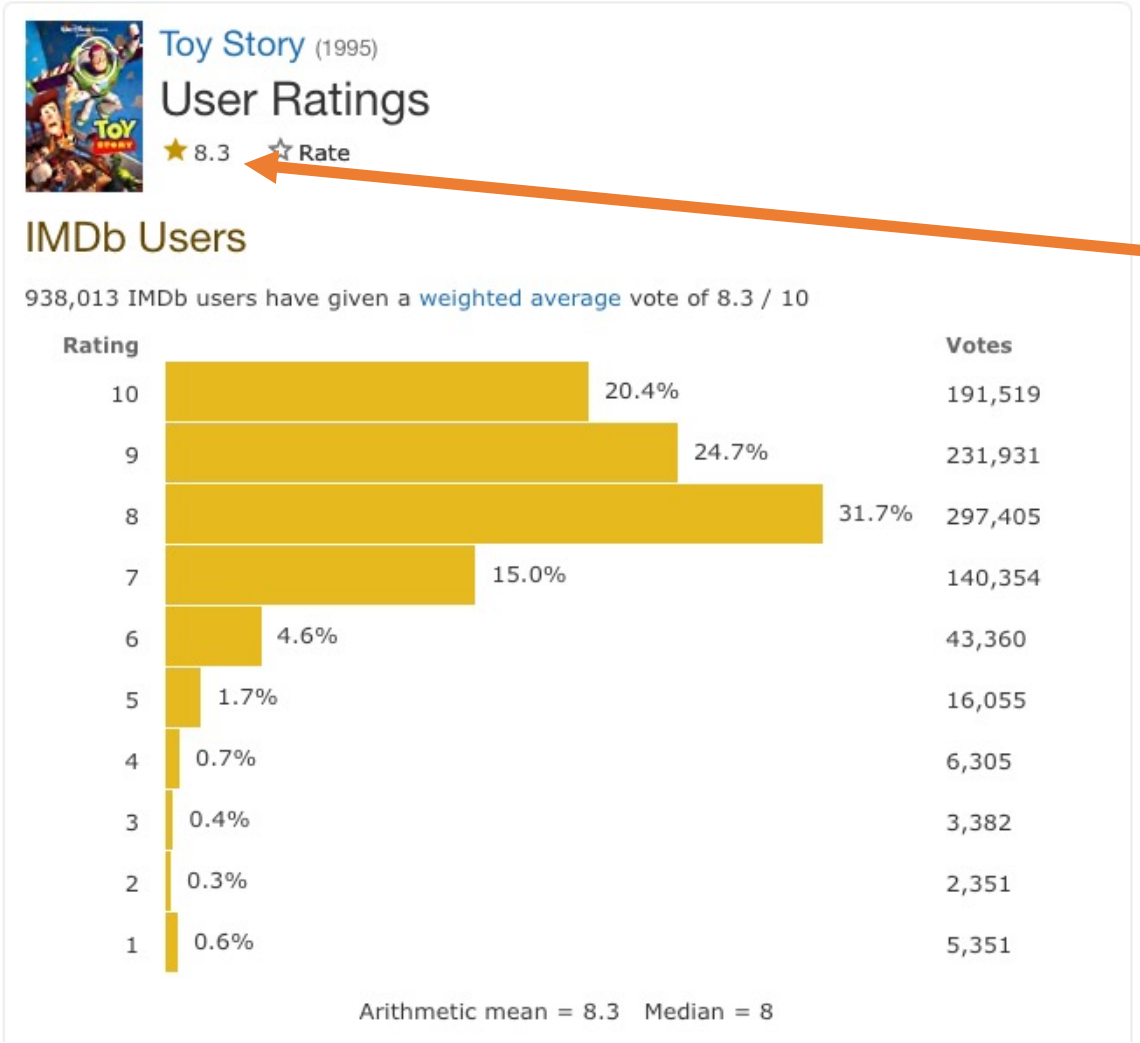
Original Classification: Adventure, Family

"Phil the Alien is one of those quirky films where the humour is based around the oddness of everything rather than actual punchlines. At first it was very odd and pretty funny but as the movie progressed I didn\'t find the jokes or oddness funny anymore. Its a low budget film (thats never a problem in itself), **there were some pretty interesting characters, but eventually I just lost interest**. I imagine this film would appeal to a stoner who is currently partaking. For something similar but better try "Brother from another planet"

Predictions: Positive 51%

The review suggests that there are some positive elements, but as they explain longer, it is clear the have negative sentiment. As a result, our model falsely predicted positive.

Sentiment Analysis Business Use Cases



Improved Rating System

As a 1-10 scale can often be open to a reviewer's interpretation, positive and negative sentiment scoring can become a new, robust addition to the rating system.

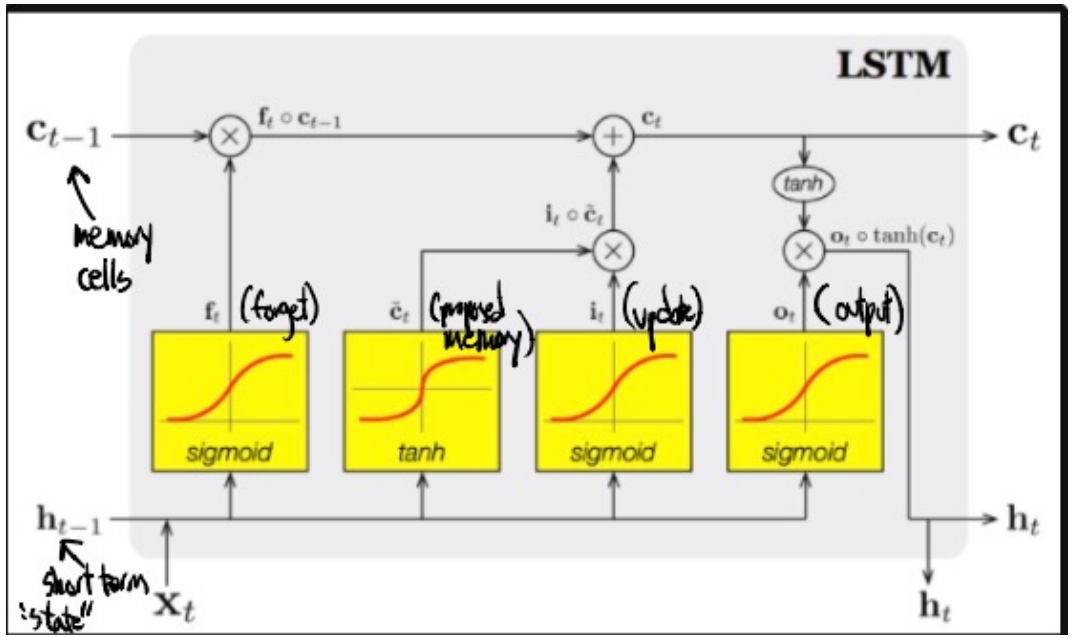
This sentiments can be factored into the weighted average or can create a new type of rating, like the rotten/fresh system at Rotten Tomatoes.

Automated Reporting

Major production companies will often want feedback on the customer sentiment of their movies.

Our model outputs can become a tool for automating some reports for these companies, delivering a valuable new data product with unique insights directly from customer feedback.

Genre Overview Model Architecture



Voacb Size: 10,000

Layers: Embedding (Using Trained Word2Vec Model), Masking, LSTM (32 units), Dense (16 Units), Dense (7 Units SoftMax)

Epochs: 5

Model Architecture:

This model was composed using our own trained Word2Vec 300 vector embedding with a LSTM layer. Long-term short memory is a special kind of RNN architecture that are designed to remember information for long periods of a time.

The model outputs the probability that the written summary belongs to a specific genre for each of the seven genre classes (Action/Adventure, Animation/Family, Comedy, Documentary, Drama, History/War/Western, Thriller/Horror)

Model Hyperparameters Tuning:

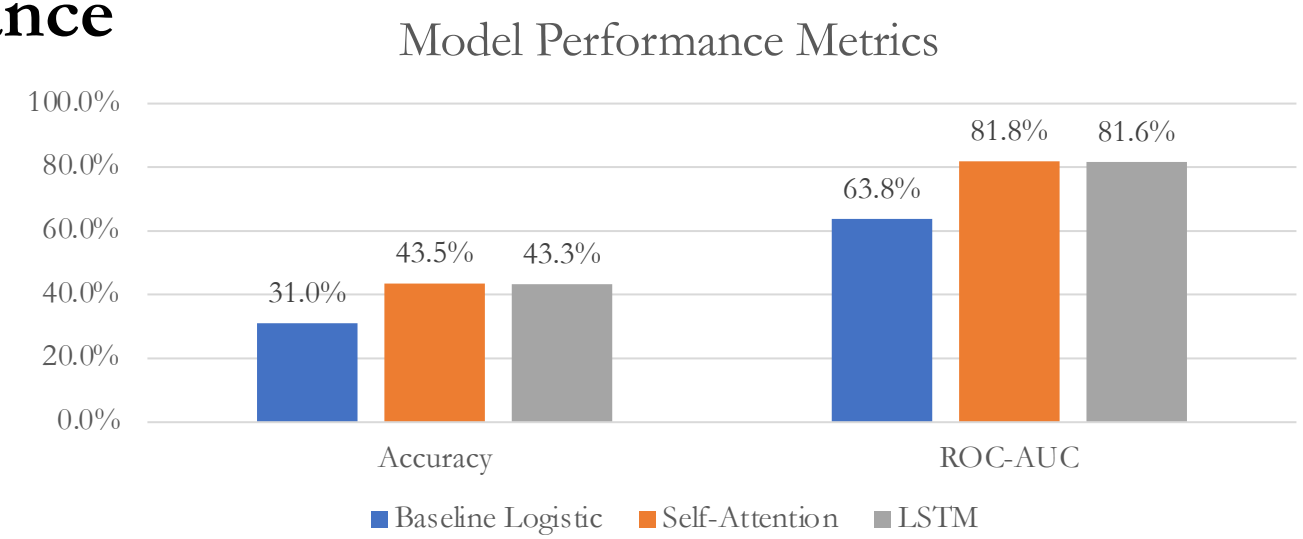
Number of Layers and Type (LTSM/Bidirectional/Attention/Dense), Epochs, Document Length, Word Embeddings (Trained vs.. pre-trained)

Appendix has full-table of tested models

Source: Recurrent Neural Networks and LSTMs for NLP Course Notes, https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Genre Overview Model Performance

- Deep Learning Architectures improved accuracy by **12 percentage points**
- Selected the base LSTM model as it performed as well as our self-attention model with less latency in training
- Our model strongly predicted the Documentary, Thriller and Horror, and Drama classes
- Although model accuracy for 7 classes is above baseline, lower accuracy performance can be attributed to many movies being tagged to multiple genres
- We recommend setting a score cutoff that can accommodate these movie summaries than span across multiple genres



Genre	Precision	Recall
Action/Adventure	0.37	0.24
Animation/Family	0.36	0.21
Comedy	0.42	0.35
Documentary	0.58	0.68
Drama	0.47	0.56
History/War/Western	0.35	0.24
Thriller/Horror	0.40	0.53

Genre Overview Model Output

Correct Prediction: Toy Story

Original Classification: Animation, Family, Comedy

"Led by Woody, Andy's toys live happily in his room until Andy's birthday brings Buzz Lightyear onto the scene. Afraid of losing his place in Andy's heart, Woody plots against Buzz. But when circumstances separate Buzz and Woody from their owner, the duo eventually learns to put aside their differences."

Predictions:

Comedy (49%) and Animation/Family (34%)

By setting a cutoff at 20%, we could correctly predict both genres for the movie Toy Story. A full prediction table can be found in the appendix

Incorrect Prediction: Jumanji

Original Classification: Adventure, Family

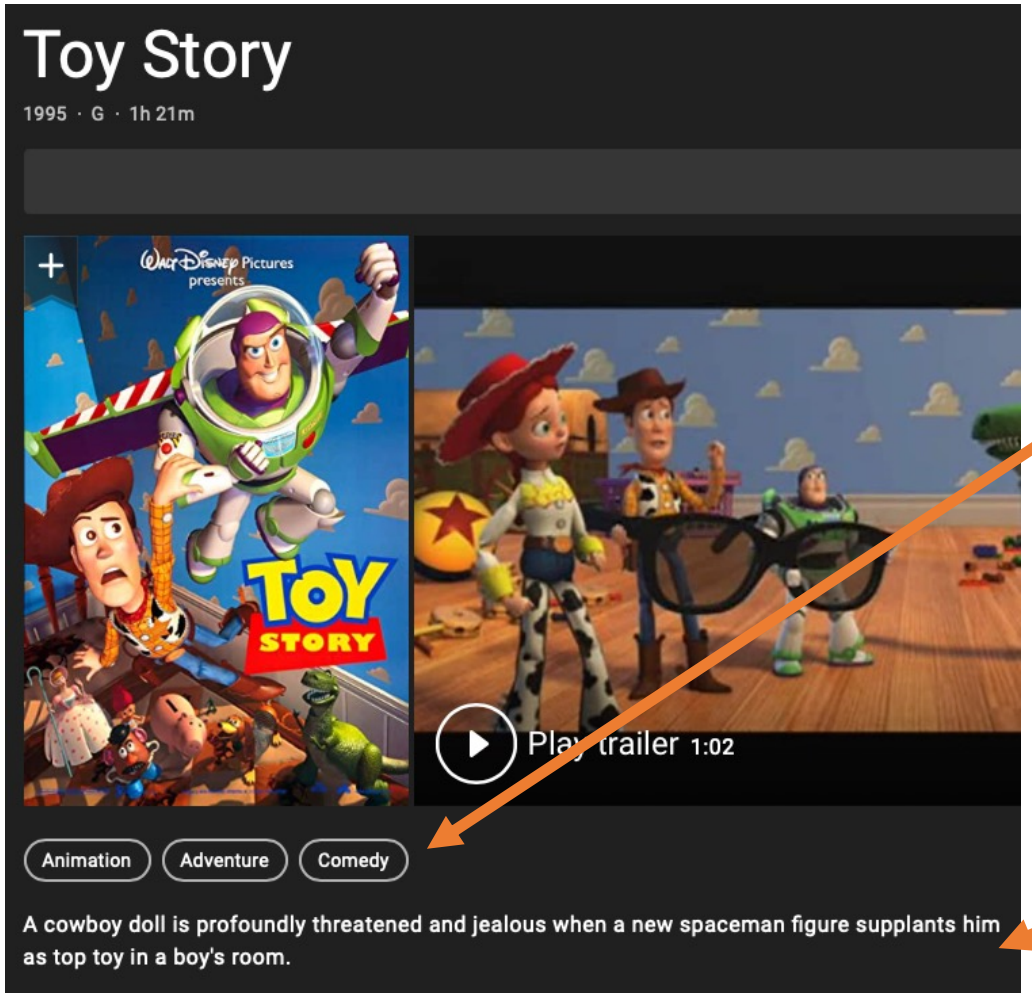
"When siblings Judy and Peter discover an enchanted board game that opens the door to a magical world, they unwittingly invite Alan -- an adult who's been trapped inside the game for 26 years -- into their living room. Alan's only hope for freedom is to finish the game, which proves risky as all three find themselves running from giant rhinoceroses, evil monkeys and other terrifying creatures"

Predictions:

Thriller/Horror (43%), Adventure (24%), Comedy (17%)

By reading the summary we can see common words that might be associated with horror/thriller films. Therefore, these misclassifications can be used as a tool to inform companies that their summary could be sending the wrong audience message.

Genre Overview Model Business Use Cases



Automatic Tagging

When a user enters the movie overview, the top matching genres would appear as selection.

Additionally, when a user pre-selects genres, movies that don't match according to the model can be flagged for further review to improve genre tagging accuracy.

Overview Feedback

For major production companies, our model outputs can generate a score to determine if an overview aligns with the intended movies genres to ensure these companies are targeting the right audience.

If scores don't match up with the movies intended genres, movie overview edits can be suggested.

Review Classification Model - Data Preprocessing and Feature Engineering

1

Due to the limited computation power, we reduced the size of training data by random sampling around 20% of records from IMDB 320,000 Movie Reviews

2

Combined the information from 3 text columns as model input, including review titles, reviews and movie titles

3

Used Regex to remove stop words detected during exploratory data analysis

4

Used pre-trained word embedding from Spacy to vectorize text corpus, obtained 900 features (300 for each text columns

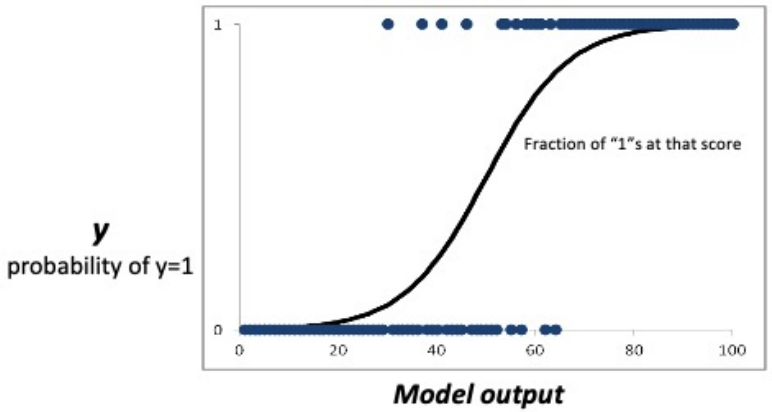
5

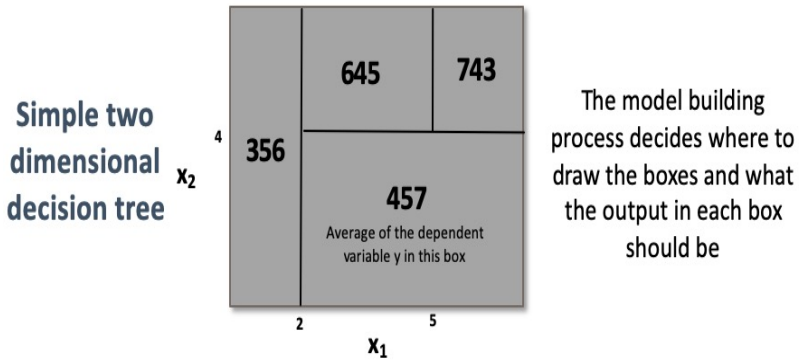
For minor genres, we categorize them as "Others" to reduce the number of targets.

6

Combining other relevant information, we got totally 906 features (900 vectors + 6 rating-related features) and a target variable of 6 categories.

Review Classification Model Architecture





The model building process decides where to draw the boxes and what the output in each box should be

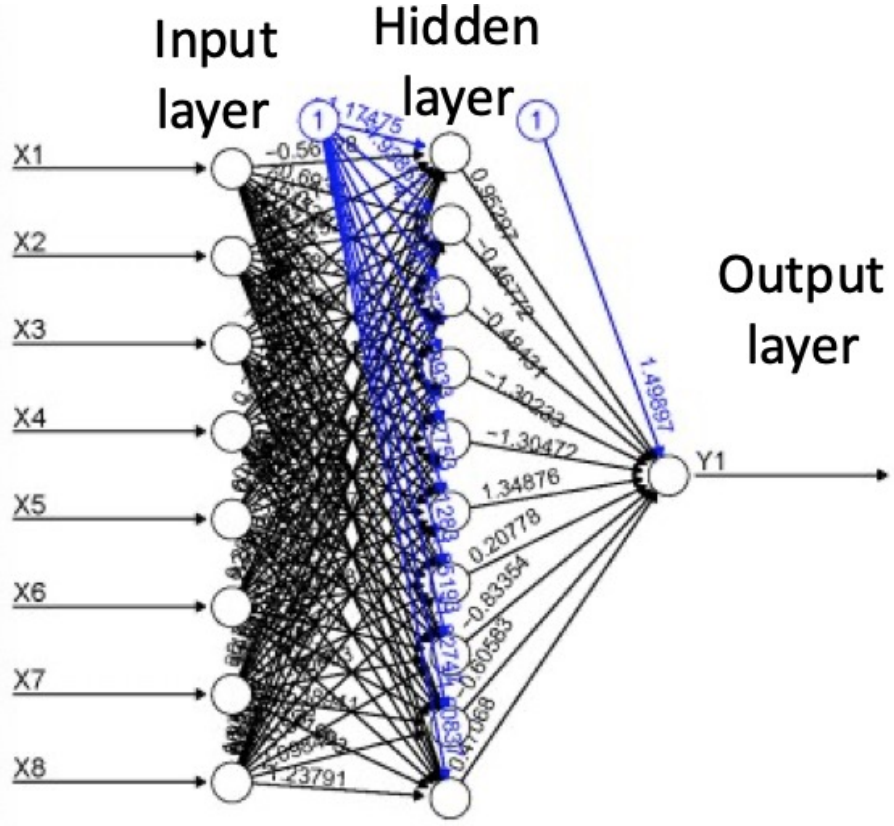
Logistic Model:

Logistic regression is a statistical model that in its basic form uses a logistic function to (usually) model a binary dependent variable (i.e., fraud label). Logistic regression is estimating the parameters of logistic model given all the X parameters for binary classification problem with logit function.

Gradient Boosted Trees:

Gradient Boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. Decision tree divides the independent variable (x's) space into boxes and places a step above each box at the height of the average of the dependent variable y in that box.

Review Classification Model Architecture



Random Forest:

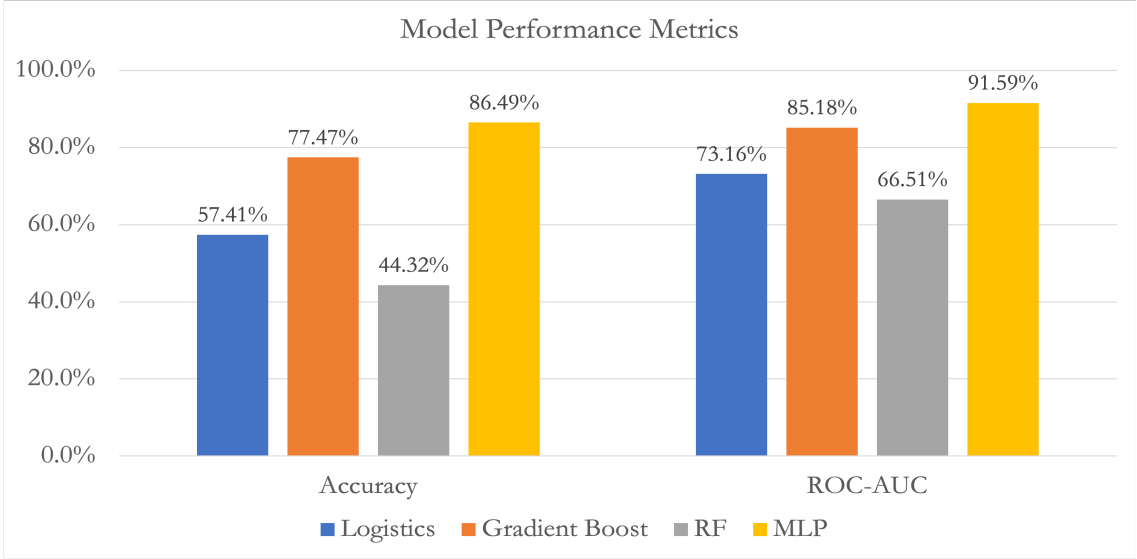
Random forest is an ensemble learning method for classification that operates by constructing multitude of decision trees. We used random forest to detect possible non-linear boundary in the feature space. One advantage of random forest is the ensembles of the model. It created many trees and then output the mode of classes for different tree algorithms.

Neural Network:

A neural network is a machine learning algorithm construct that maps an input vector to an output scalar, or typically a vector of axes into a single dependent variable. It is inspired by the biological neural networks that constitute human brains. A typical neural network consists of an input layer, some hidden layers and an output layer.

Review Classification – Base Model Performance

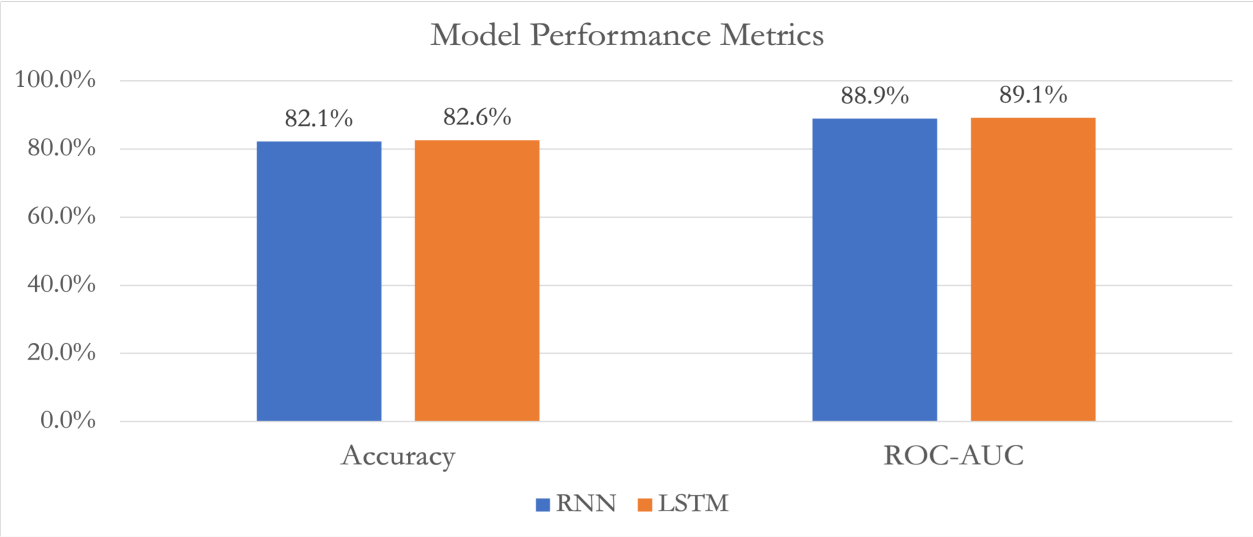
- Our base models include Logistic Regression model, Gradient Boosting classifier, Random Forest Classifier, and Multi-layer Perceptron classifier
- Logistic Regression is the fastest base model but provides poor results
- Gradient Boosting and Random Forest Classifier did poor jobs predicting correct genres and generally took long time to run
- MLP Classifier is the best choice in this case, it is relatively fast and gives the greatest prediction accuracy
- The scores for all the 6 classes are relatively balanced, with Drama and Others slightly lower than the other classes.



Genre	Precision	Recall
Comedy	0.86	0.87
Action	0.81	0.94
Drama	0.91	0.69
Others	0.78	0.83
Crime	0.82	0.81
Animation	0.89	0.82

Review Classification – Deep Learning Model Performance

- Self-trained word embedding using the movie review columns, which improved the model performance significantly (refer to appendix for model performance).
- LSTM model performs slightly better than the RNN model, but still worse than the MLP baseline model
- Generally, our model has an accurate prediction for all the 6 genres.
- The scores for all the 6 classes are relatively balanced, with Drama and Others slightly lower than the other classes.



Genre	Precision	Recall
Comedy	0.85	0.85
Action	0.84	0.89
Drama	0.77	0.81
Others	0.83	0.75
Crime	0.81	0.79
Animation	0.90	0.82

Model Performance Evaluation

Logistic

In-Sample Accuracy:

58.39%

Out-of-sample Accuracy:

57.41%

Average runtime:

3 mins

MLP

In-Sample Accuracy:

99.33%

Out-of-sample Accuracy:

86.48%

Average runtime:

5 mins

RNN

In-Sample Accuracy:

90.52%

Out-of-sample Accuracy:

82.13%

Average runtime:

52 mins

LSTM

In-Sample Accuracy:

95.56%

Out-of-sample Accuracy:

82.58%

Average runtime:

43 mins

Review Classification Model Output

Correct Prediction: A Life Less Ordinary

Original Classification: Comedy

"I'm a big fan of Cameron Diaz, and Ewan McGregor caught my attention after his riveting performance in Trainspotting. McGregor is the bumbling kidnapper and Diaz is the victim who seems to know more about kidnapping than he does. All this makes for a really funny comic act from the duo as they progress from kidnapper/victim to lovers..."

Predictions:

Comedy

From the review, we can see that there are words like "funny" and "comic" that clearly indicated that it is a comedy.

Incorrect Prediction: Reggie's Prayer

Original Classification: Drama

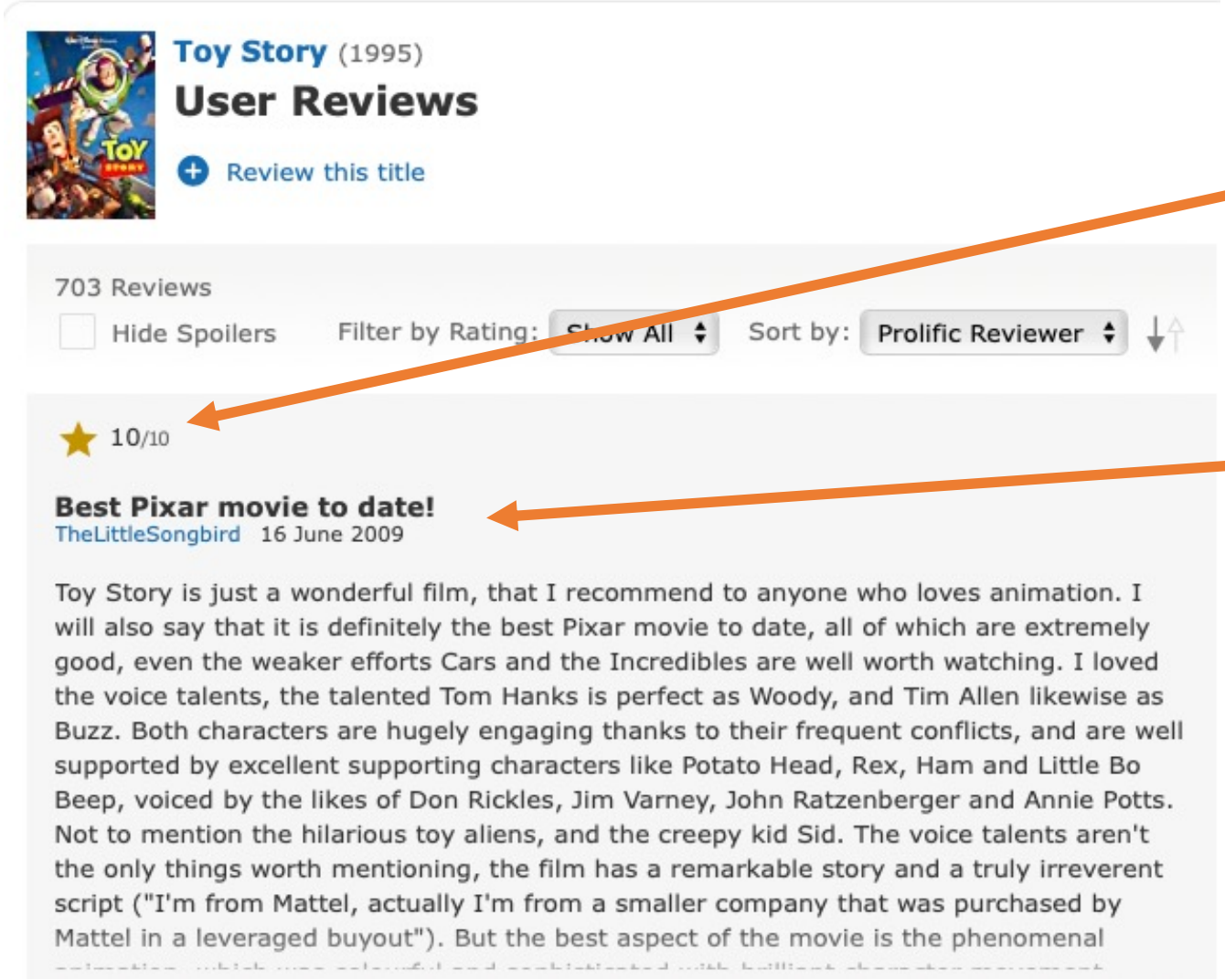
"A player-turned-coach story with the main character being played by a retired football player gives this film an authentic edge that makes the story line come to life. It's a must-watch for any football fan."

Predictions:

Action

By reading the summary we can see the movie is related to football and there are words like "player" that might be associated with action movie.

Review Classifier Business Use Cases



Score Re-Weighting

As each user my have different criteria, we can use a review quality rating to reweight the 1-10 scores. This can lead to a more precise ranking system that is influenced by more quality reviews.

Review Quality Rating

Since our model can strongly predict the genres that a movie are associated with, we can use these predictions to determine the quality of a movie rating.

For example, if a single review simply said, "best movie ever", our model is likely to not determine which genre belongs to that review.

As a result, we can quickly find higher quality reviews and promote as higher quality content on the site.

Implementation Roadmap

Latency and Scaling

For production, all text preprocessing models should be pre-trained, so they integrate seamlessly into a feature engineering pipeline.

This includes creating production objects that performs regex cleaning, removes stop words, lemmatizes the data, gets document embeddings, and converts to a padded doc structure.

We successfully scored 65,000 records in 21 seconds for the genre overview model and scored 50,000 seconds in 45 seconds for the sentiment analysis model. This can be scaled for real-time scoring.

Recommendations

One of our implementation recommendations is to have new genre tag options appear when a new movie summary is written.

Leveraging the model outputs, scores can be rank ordered, with the top genres appearing first for the user to quickly select as a genre classification.

Our sentiment analysis model can auto generate reports that can be used as a basis for future recommendations for customer response to movies

Model Detrition and Drift

Since our model poorly predicts the Animation/Family/History/War/Western genres, we will want to improve the model in the future.

We recommend collecting more data on these classes and rebuilding the model with these new observations.

Additionally, our production pipeline should include model monitoring to track word frequency data drift and performance metrics overtime.

ROI: Sentiment Analysis Model

Add another rating system that can compete with other review aggregators

Give insights on the customer sentiment and segments. Opens new revenue/consulting opportunities to film companies

Earnings and Savings (Monthly)*

Rating System – $375K (New Traffic Redirected from other Review Aggregators, 5 million new users), Automated Sentiment Reports– $250K (On ad-hoc basis, up to 2000 reports a month, saving 5,000 minutes per month), Consulting - $400K (Sentiment Reports for Production Companies)

Expenditures (Monthly)

Development (Fixed) $50K (1 months on team of data scientists and ML engineers), Engineering Requirements - $20K, Technical Costs $50K (Implementation and Testing), Misc - $20K

Return on Investment

(Earnings and Savings Potential/$1.025M * 12 – Expenditures/$90K * 12 – $50K) = **$10.72 Million**

*Assumes $15 CPM, 5 impressions per unique user, $50 each hour saved (Data Scientist Salary)

ROI: Review Classification Model

Determine quality of movies reviews to create a new dimension that feeds into the rating system

Earnings and Savings (Monthly)*

Improved Rating System – $375K (5% increase in impressions for full userbase), Increased Traffic from Promoting Strong Reviews – $169K (15% increase in impressions from dedicated user base – 5 million users)

Expenditures (Monthly)

Development (Fixed) $150K (3 months on team of data scientists and ML engineers), Engineering Requirements - $100K, Technical Costs $75K (Implementation and Testing), Misc - $35K

Return on Investment

(Earnings and Savings Potential/$544K * 12 – Expenditures/$210K * 12 – $150K) = **$3.85 Million**

*Assumes $15 CPM, 5 impressions per unique user, 15 impressions for dedicated user

ROI: Genre Overview Model

Automate the genre tagging process for movie catalog.
Improves efficiency and creates new traffic potential

Give feedback on the effectiveness of summaries.
Opens new revenue/consulting opportunities to film companies

Earnings and Savings (Monthly)*

Product – $180K (Users can score their own custom summaries leading to 10% increase in impressions for 8M users), Automated Tagging Verification – $60K (2K new movies/shows added each month, saving 4,000 minutes per month), Consulting - $250K (Data Product for Production Companies)

Expenditures (Monthly)

Development (Fixed) $150K (2 months on team of data scientists and ML engineers), Engineering Requirements - $50K, Technical Costs $50K (Implementation and Testing), Misc - $20K

Return on Investment

(Earnings and Savings Potential/$490K * 12 – Expenditures/$120K * 12 – $150K) = **$4.29 Million**

*Assumes $15 CPM, 15 impressions per unique user, $15 each hour saved (Data Entry Salary)

Final Return on Investment

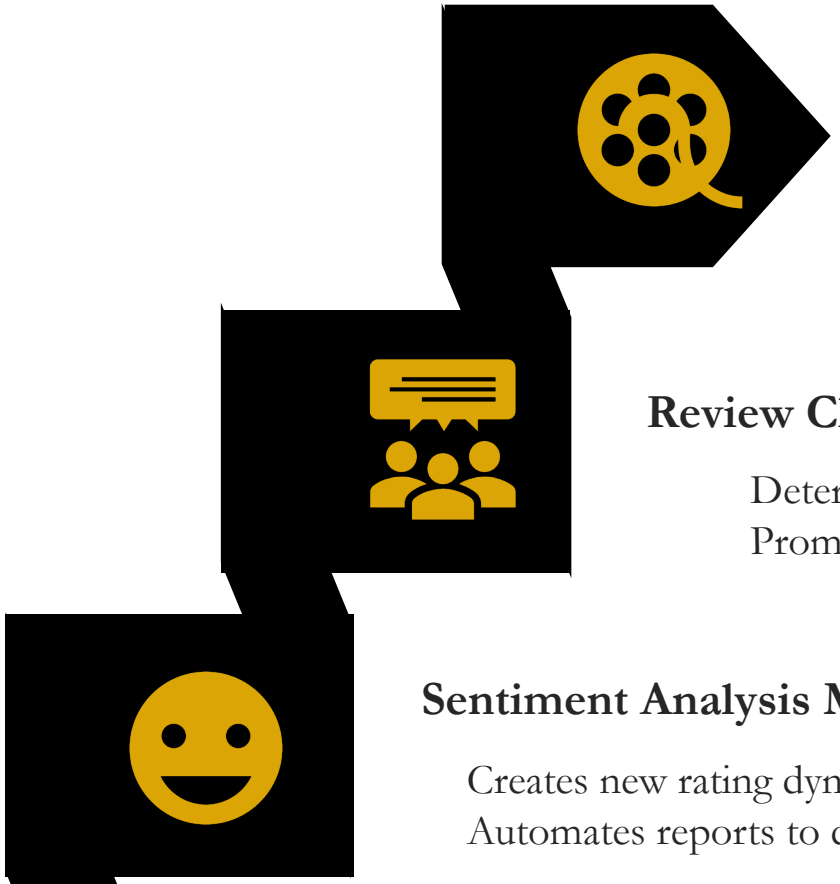
Sentiment Analysis Model	Review Classification Model	Genre Overview Model
$8 – 15 Million	$2.5 – 4.5 Million	$2.5 – 5 Million

Return on Investment

$13 - $25M

Conclusion

- Advanced NLP techniques can create new business opportunities and improve operational efficiency
- By packaging model outputs into products, IMDB can offer advanced analytical services to production companies creating new revenue streams
- These models can also influence a new rating system, driving more traffic



Genre Overview Model

Automate the genre tagging process
Provide feedback on movie summaries

Review Classification Model

Determine review quality and integrate into ratings
Promote quality reviews to front of page

Sentiment Analysis Model

Creates new rating dynamic for users
Automates reports to deliver to companies

Improved Operational Efficiency
$18M Return on Investment

Thank You

IMDb

Appendix

Sentiment Analysis Model Hyperparatmers Tests

Architecture	Max Doc Length	Epochs	Vocab Size	Max Features	Accuracy / ROC AUC
Logistic Regression with CountVectorizer	N/A	N/A	N/A	3000	87.34%/0.8734
Support Vector Classification with CountVectorizer	N/A	N/A	N/A	3000	87.27%/0.8727
Multinomial Naive Bayes with CountVectorizer	N/A	N/A	N/A	3000	84.17%/0.8417
Logistic Regression with TfidfVectorizer	N/A	N/A	N/A	3000	88.06%/0.8806
Support Vector Classification with TfidfVectorizer	N/A	N/A	N/A	3000	87.92%/0.8792
Multinomial Naive Bayes with CountVectorizer	N/A	N/A	N/A	3000	84.82%/0.8482
MLPClassifier(max_iter=300)	NA	NA	NA	NA	86.1%/0.861

Sentiment Analysis Model Hyperparatmers Tests (Cont.)

Architecture	Max Doc Length	Epochs	Vocab Size	Word Embeddings	Accuracy / ROC AUC
Sequential 3 Layers (Embedding, Flatten, Dense)	400	5	10000	300	88.5%/0.885
RNN 4 layers (Embedding, Masking, SimpleRNN(64), Dense (16))	300	5	10000	Trained Word2Vec (300,)	74.2%/0.741
LTSM 4 layers (Embedding, Masking, LSTM(64), Dense (16))	600	5	10000	Trained Word2Vec (100,)	89.7%/0.897
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	300	10	10000	Trained Word2Vec (300,)	88.6%/0.886
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	400	5	5000	Trained Word2Vec (300,)	88.5%/0.885
LTSM 4 layers (Embedding, Masking, LSTM(64), Dense (16))	**300**	**5**	**10000**	**Trained Word2Vec (300,)**	**90.3%/ 0.903**
LTSM 4 layers (Embedding, Masking, LSTM(64), Dense (16))	1400	5	10000	Pretrained Glove (100,)	89.1%/ 0.891
Self Attention 5 layers (Embedding, Masking, Self Attention, LSTM(32), Dense (16))	300	5	10000	Trained Word2Vec (300,)	89.1% / 0.891

Review Classification Model Hyperparatmers Tests

Architecture	Penalty	Solver	Multi_class	Word Embeddings	Accuracy
Logistic Regression	l2	lbfgs	auto	Pre-trained word2vec	68.20%

Architecture	n_estimator	max_depth	learning rate	Word Embeddings	Accuracy
Gradient Boosting	100	3	0.1	Pre-trained word2vec	77.47%

Architecture	n_estimator	max_depth	criterion	Word Embeddings	Accuracy
Random Forest	300	3	entropy	Pre-trained word2vec	44.32%

Review Classification Model Hyperparatmers Tests

Architecture	Hidden Layer Size	Activation	Solver	Word Embeddings	Accuracy
MLP	(100,)	relu	adam	Pre-trained word2vec	86.46%
MLP	(50,50,50)	tanh	adam	Pre-trained word2vec	84.74%
MLP	(50,100,50)	tanh	adam	Pre-trained word2vec	85.02%
MLP	(150,50,50)	tanh	adam	Pre-trained word2vec	85.34%
MLP	**(150,100,150)**	**tanh**	**adam**	**Pre-trained word2vec**	**86.48%**
MLP	(150,150,150)	tanh	adam	Pre-trained word2vec	83.68%

Review Classification Model Hyperparatmers Tests

Architecture	Max Doc Length	Epochs	Vocab Size	Word Embeddings	Accuracy
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	410	5	136250	Pre-trained GLOVE (100,)	68.20%
RNN 4 layers (Embedding, Masking, LSTM(64), Dense (16))	410	5	136250	Pre-trained GLOVE (100,)	59.60%
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	676	5	136250	Trained Word2Vec (300,)	77.70%
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	**410**	**3**	**136250**	**Trained Word2Vec (300,)**	**82.58%**
RNN 4 layers (Embedding, Masking, LSTM(64), Dense (16))	506	5	136250	Trained Word2Vec (**300**,)	71.82%
RNN 4 layers (Embedding, Masking, LSTM(64), Dense (16))	410	3	136250	Trained Word2Vec (300,)	82.14%

Genre Overview Model Hyperparatmers Tests

Architecture	Max Doc Length	Epochs	Vocab Size	Word Embeddings	Accuracy / ROC AUC
Logistic Regression with Word Embeddings (Baseline)	N/A	N/A	N/A	Pre-trained (96,)	31.04% / 0.638
LTSM 4 layers (Embedding, Masking, LSTM(64), Dense (16))	146	5	10000	Pre-trained GLOVE (100,)	42.0% / 0.817
LTSM 4 layers (Embedding, Masking, LSTM(64), Dense (32))	146	10	10000	Pre-trained GLOVE (100,)	39.8% / 0.808
LTSM 4 layers (Embedding, Masking, LSTM(64), Dense (16))	**146**	**5**	**10000**	**Trained Word2Vec (300,)**	**43.3% / 0.816**
LTSM 4 layers (Embedding, Masking, LSTM(64), Dense (16))	146	5	10000	Trained Word2Vec (100,)	42.2% / 0.814
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	146	5	10000	Trained Word2Vec (100,)	42.2% / 0.814
LTSM 4 layers (Embedding, Masking, LSTM(128), Dense (16))	146	5	10000	Trained Word2Vec (100,)	42.1% / 0.815
LTSM 4 layers (Embedding, Masking, LSTM(128), Dense (16))	146	5	20000	Trained Word2Vec (100,)	42.8% / 0.816

Genre Overview Model Hyperparatmers Tests (Cont.)

Architecture	Max Doc Length	Epochs	Vocab Size	Word Embeddings	Accuracy / ROC AUC
Bidirectional LSTM (Bi LSTM(300), LSTM (128)	146	5	10000	Embedding layer (100,)	40.3% / 0.809
Self Attention 5 layers (Embedding, Masking, Self Attention, LSTM(128), Dense)	146	5	10000	Trained Word2Vec (100,)	43.4 % / 0.818
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	146	20	10000	Trained Word2Vec (100,)	42.3% / 0.813
Self Attention 5 layers (Embedding, Masking, Self Attention, LSTM(32), Dense (16))	146	10	10000	Trained Word2Vec (100,)	40.34% / 0.809
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	146	20	10000	Trained Word2Vec (100,)	42.5% / 0.813
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	120	10	15000	Trained Word2Vec (300,)	42.6% / 0.820
LTSM 4 layers (Embedding, Masking, LSTM(32), Dense (16))	100	10	10000	Trained Word2Vec (300,)	42.4% / 0.815

Genre Overview Model Full Predictions

Genre	Toy Story	Jumanji
Action/Adventure	9.9%	**24.1%**
Animation/Family	**34.2%**	9.4%
Comedy	**49.3%**	**16.7%**
Documentary	0.1%	0.2%
Drama	5.6%	6.8%
History/War/Western	0.1%	0.0%
Thriller/Horror	0.7%	**42.9%**