# What is MST3K and How is it Affecting my IMDb Review Prediction Model?

During one of my lectures on machine learning we were coving how to develop a model that could predict if an IMDb user movie review was positive or negative. The basic premise of the model is it looks at each word in the review and then whether each word has a positive or negative word associated and then it takes a sum of those weights. If the total sum of weights is positive, then the model would predict a positive review and vice versa for negative. The further the sum of the weights is away from zero the more confident the model is about its prediction.

I was curious about what words the model considered positive and negative words. So, I made the following table:



Most of these words made sense due to their negative connotation. I would expect words like “disappointment” and “forgettable” to be good predictors of negative reviews: however, the 3rd most negative “**mst3k**” word is the most interesting.

## What is MST3K?

MST3K or Mystery Science Theater 3000 was a television show that had over 200 episodes spanning from 1989 to 1999. So, why would the shorthand for a television show from the 80s have an influence on our model? To answer that, we have to first understand the premise and the cultural impact that this show had. The premise of the show is that the star, Joel Robinson, is trapped in space with only B movies to watch. So, he and his two robot friends sit-down and watch these movies while providing their own commentary as the movie is playing.

A picture containing person

Description automatically generated

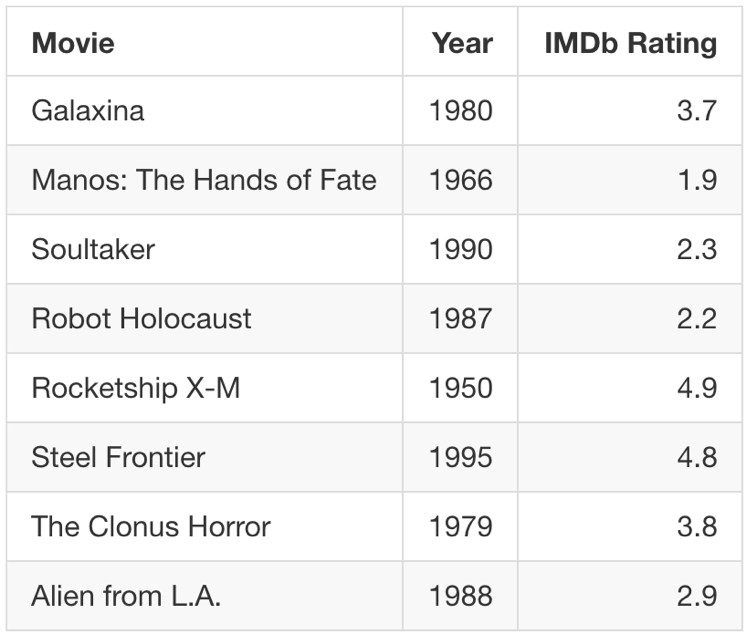
*Source: S3:E8 “Deathstalker and Warriors from Hell”*

The appeal of the show comes from hearing the series mains discuss the continuity mistakes and running gags in real-time as the terrible film plays in the background. It is reminiscent of a movie night with your friends as you crack a couple of beers and everyone shouts over the movie with their own jokes. This style of critiquing films became a highly adapted style when YouTube was in its infancy. You can still see echoes of this style in popular YouTube channels such as RedLetterMedia’s *Half in the Bag and Best of the Worst*, and Screen Junkies’ *Honest Trailers*.

## So how does this relate to the model?

We have established that MST3K was a cultural hit with cinephiles. With this much wide appeal it stands to reason that some fans would hunt down the original movies featured on MST3K give them watch and review them afterwards. If we read through some of these reviews, we can find what types of movies these were by searching them up on IMDb.

The following are a few of the movies that were reviewed and contained the phase “MST3K” in the IMDb dataset:



*Source: IMDb*

As you can see, the majority of these films have negative review scores, which makes sense considering these B-movies were selected to appear on MST3K to begin with. Therefore, our model is using “MST3K” as a proxy for B-movie. The issue with this is MST3K only applies to some reviews in our dataset and does not have the mass generalization a word such as “laughable” and “forgettable” would have.

We can see this issue if we look at how often “MST3K” was used compared to “Waste”:

A screenshot of a cell phone

Description automatically generated

As you can see above, “MST3K” proportionally is a better predictor of a negative review than “Waste”; however, “Waste” is present in 10x the number of the reviews than “MST3K”.

## How do we fix this?

One way we could tackle this issue could be to use a **TFIDF** or **term frequency–inverse document frequency** model. This method relies on the weight value of how often a word appears in all the reviews. Therefore, words such as “Waste” will have a higher weight in the model.

The following is list of the top weighted negative word from the TFIDF model:

A screenshot of a cell phone

Description automatically generated

This model seems to have weighted words that are more common in reviews than more uncommon words such as MST3K. Additionally, it decreased the validation error (how well the model works on untrained on data) from the original 13% error rate to an 11% error rate of TFIDF model. That is a free 2% decrease in error just for changing the type of our model.

## Conclusion

As a data scientist, it is important to look at the interoperability of our models and question why our model is predicting something. The more you understand data and how it is interacting with your model the more, the better we can do to improve our model. If you are curious about to how to build models such as the one above, feel free to look at the model code (model)[ <https://github.com/MrThomasPin/what-is-mst3k/blob/master/imdb_mst3k.ipynb>].

(Main GitHub repo) [ <https://github.com/MrThomasPin/what-is-mst3k>]

(Model) [<https://github.com/MrThomasPin/what-is-mst3k/blob/master/imdb_mst3k.ipynb>]

(Oscar prediction model)[ <https://www.linkedin.com/pulse/oscar-machine-learning-model-vs-data-scientists-girlfriend-thomas-pin/>]

By Thomas Pin

Edited by Sarah Gallagher