Deterministic Machine Learning: Feature Importance of Persuasive Linguistic Predictors

Leonardo Carrico

Follow this and additional works at: https://publications.lakeforest.edu/seniortheses

Part of the Computer Sciences Commons

Recommended Citation
Deterministic Machine Learning: Feature Importance of Persuasive Linguistic Predictors

Abstract
Generally, people's opinions are difficult to change. This study searches for the most important features of persuasion. Identifying social media as one of the most difficult modes through which to change people's opinions, we discovered a subreddit dedicated to the sole purpose of changing opinions. Using the data set obtained from r/ChangeMyView, we identified opinion-changing comments and ran machine learning algorithms on those comments. We then further explore those machine learning models to identify how exactly they work. By determining the feature importance and weights of the models, we will specific aspects of persuasive linguistics that influence people's original opinions.

Document Type
Thesis

Degree Name
Bachelor of Arts (BA)

Department or Program
Data Science, Math, and Computer Science

First Advisor
Arthur Bousquet

Second Advisor
Vivian Ta

Third Advisor
Enrique Treviño

Subject Categories
Computer Sciences

This thesis is available at Lake Forest College Publications: https://publications.lakeforest.edu/seniortheses/161
Deterministic Machine Learning:
Feature Importance of Persuasive Linguistic Predictors

by

Leonardo Carrico

April 24, 2020

The report of the investigation undertaken as a
Senior Thesis, to carry one course of credit in
the Department of Data Science, Math, and Computer Science

Davis Schneiderman
Krebs Provost and Dean of the Faculty

Arthur Bousquet, Chairperson

Vivian Ta

Enrique Treviño
Abstract

Generally, people's opinions are difficult to change. This study searches for the most important features of persuasion. Identifying social media as one of the most difficult modes through which to change people's opinions, we discovered a subreddit dedicated to the sole purpose of changing opinions. Using the data set obtained from r/ChangeMyView, we identified opinion-changing comments and ran machine learning algorithms on those comments. We then further explore those machine learning models to identify how exactly they work. By determining the feature importance and weights of the models, we will specific aspects of persuasive linguistics that influence people's original opinions.
Acknowledgments

I gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU to our research lab used for this research.

I would like to acknowledge my professors, Arthur Bousquet, Vivian Ta, and Enrique Treviño for their knowledge and support. Thank you, Prof. Bousquet, for teaching me fundamental concepts around Machine Learning to be able to build my skills in this field as well as the support needed to continue writing this thesis. Thank you, Prof. Ta, for the help obtaining the dataset, as well as the prior research needed to generate the necessary features for the Machine Learning Algorithms. Thank you, Prof. Treviño, for the kindness, the contributions for the thesis, and the adorable, keyboard-stealing daughter.

I would also like to thank Liam Connell for the opportunity to begin my research over the summer on this topic, as well as his ideas to add on to the research.

Lastly, I would like to thank my cat, Oreo, for meowing at me whenever I would get distracted with other activities.
# Table of Contents

1 Introduction .......................................................................................................................... 1
   1.1 Libraries/Modules .......................................................................................................... 4
2 Dataset .................................................................................................................................. 6
   2.1 Cleaning Data ................................................................................................................ 6
   2.2 Balancing Data .............................................................................................................. 7
3 Feature Engineering ............................................................................................................. 9
   3.1 Features based on NLP techniques .............................................................................. 9
   3.2 Features based on Persuasive Linguistics .................................................................. 12
4 Results .................................................................................................................................. 14
   4.1 Models .......................................................................................................................... 14
   4.2 Accuracy Score/Confusion Matrix .............................................................................. 15
   4.3 Feature Importance ..................................................................................................... 16
5 Text Generation .................................................................................................................. 19
6 Conclusion ............................................................................................................................ 25
References ............................................................................................................................... 27
1 Introduction

Practically everyone has the desire to change another person’s opinion or view on a topic. This goes from political campaigns, to general conversations, even to deciding what restaurant to go out to with your significant other. From all these areas and categories, conversation and persuasion have always been goals and skills that people reach for and develop. However, many of these acts of persuasion tend to fail or become completely misunderstood. Some of these attempts are even seen as attempts to anger the other party.

A lot of these persuasive attempts are seen on social media, where one person states their opinion online, and then others reply and comment their own. The purpose of the post can vary depending on the person and their intentions. There are some who post so others can learn about their opinions, and some who post in order to listen and understand the opinions of others. Reddit, a social media platform for forums, has a subreddit or a specific forum, only designated for posting your opinion so others can attempt to change it. This subreddit is named /r/ChangeMyView, also known as CMV for short. Compared to other forms of social media, since CMV is specific for posting opinions, they have rules that need to be followed in order to post and comment in this forum, and administrators who regulate the forum. Due to this, forms of attack and aggression are restricted turning this forum into an online location for civil discourse.
For the conversation to begin, the post must first be made. There is a set of submission rules that needs to be followed to ensure that the post is clear, expressive, and open-minded. One purpose of these rules is to ensure that opinions are properly listened to and carefully considered. These rules give more opportunity to those commenting since their opinions are more likely to be explored and truly listened to. Typically, the original poster, also known as OP, will reply to some of the individual comments that spark interest, and may have the possibility of their opinion getting changed, see Figure 1.

![Reddit Post](https://i.imgur.com/123456.png)

**Figure 1: Example of a Submission**

If the OP’s opinion was changed by a comment, he will reply to that comment with either “!delta” or $\Delta$. This unique factor allows us to have actual data, on which arguments are most convincing on this forum.
The data that we have consists of thousands of the submissions, posts, delta-winning comments, and non-delta-winning comments. Once we have this large amount of data, we can use data science techniques to pull relationships between delta/nondelta comments with their features and determine which persuasive linguistics have the highest impact. Data science is essentially the combination of computational and mathematical methods of deriving observable and hidden knowledge from, typically, a large set of data, as explained by (Berinato, 2019).

One of the great beauties of data science is that this is an interdisciplinary field that branches to almost every section of science and even to humanities and linguistics. The
only necessary thing to be able to use this field is the data required to run its algorithms. These days, the gaining of information is tremendous which has allowed us to continuously build and grow our techniques of forming more data, based on the data that we already have. With our current data set of nondelta-winning comments and delta-winning comments, we can analyze and attempt to draw meaningful conclusions that allow us to expand our current knowledge on what impacts our attempts for persuasion and how to improve them.

Typically, the algorithms for interpreting our data follows a general pattern. We begin by obtaining our data, and then removing rows with missing or misplaced features. Once we have our data fixed up, we begin by running analysis on our data to figure out pattern and relationships between the individual predictive linguistics and their labels. Finally, we use classifiers that take in the original data to create a prediction as to whether a comment received a delta. With several classifiers working and running predictions, we can use our results to form conclusions that give us useful and efficient insight into our data.

1.1 Libraries/Modules

One great assistance to being able to program and display these machine learning algorithms are the already pre-existing libraries and modules. They allow me to use programs that other people have pre-built to quickly manipulate and run predictions on my data.
NumPy, see (Oliphant, 2006) and (van der Walt, Colbert, & Varoquaux, 2011), is one of the Python libraries used mainly for array-processing. It also contains Linear Algebra libraries that allow us to use computational methods to solve problems. We use this library in order to more easily create/generate our features from our dataset. Since it helps us with processing, we can use this module, combined with other functions, in our feature engineering process.

Another used module is called Pandas, see (McKinney, 2010). Pandas is a software module used mainly for data manipulation and analysis. It allows us to do quick operations on a dataset. Since most of our data is contained as CSV files, pandas allow us to easily import the files as datasets within our code. Once it has been imported, we can use Pandas to then convert our data into arrays to allow NumPy to process them.

The module scikit-learn, see (Pedregosa, et al., 2011), is the general module we use to generate our models. Scikit-learn has all the tools necessary for the actual data science and machine learning process and is what we use for our models. We used scikit-learn to create our Machine Learning models, which then took our array features from NumPy and turned into our prediction functions that returns a probability to whether a comment would receive a delta.

Matplotlib, see (Hunter, 2007), is used for plotting the results and doing visual analysis of our data. It allows us to quickly and efficiently visualize what our results represent and to also do preprocessing analysis on our dataset. We use the outcomes and distributions in our data to understand what has the highest impact in receiving deltas. In
order to visualize these differences, we use Matplotlib to generate plots of frequency distributions, feature importance and other useful visualizations.

2 Dataset

We first start out with obtaining then maneuvering around a dataset. A dataset is a collection of data meant to be used or processed by a computer. Datasets are made up of several parts and may be composed of several different types of values. For example, a dataset could contain the number of times a specific word appears within a post or comment, but it could also contain the words themselves. In our case, our initial dataset is composed of the text written by the authors of the comments and submissions as well as some information about the comments and submissions themselves. Though we received this data already pulled from Professor Ryan Boyd at Lancaster University in the UK, it is also possible to use PRAW, the Python Reddit API Wrapper to pull this data. PRAW is an API or set of computer functions that allows you to pull data, like our comments and posts, directly from the Reddit website. After texting the PRAW API, we found that it would be relatively easy to generate our usable dataset from CMV.

2.1 Cleaning Data

In order to use our dataset, we must first clean it. Depending where you obtained your data from, there could be a chance that the data is inaccurate, old, corrupted, or possibly unusable. Typically when going through a new dataset, you need to check to make sure the data is up-to-date, that way the results from the data would still be valid. Other cases
are when the data is inaccurate or filled with lots of human error due to manual input, or corrupted, possibly from loss of data when downloading online. When we go through the process of data cleaning, we eliminate the outlying differences and inaccuracies within the data. This can be done by deleting some of the data that is corrupt or inaccurate, but also by possibly modifying it so it can properly fit our standards. One possible method of doing this is manually, where you go through the data one by one, and try to identify specific areas where you can spot mistakes, but typically if your dataset contains thousands or even millions of data, it’s practically impossible to go through it by hand.

In our dataset, the main points of concern were removing comments and submissions that came as empty text, changing special characters like \( \frac{3}{4} \) and \( \div \) into words, and finally, removing punctuation from our text to simplify the algorithm for categorizing our words. Once our data became more accurate and usable, we could move on to handling our actual data.

### 2.2 Balancing Data

One of the first issues that we had was the difference between the number of delta-winning comments and nondelta-winning comments. With our data cleaned, we had 124,296 nondelta-winning comments and 12,138 delta-winning ones. The most initially clear effect that we can witness from this imbalanced dataset is within our attempt to use data science to predict which comments are delta-winning. Since our data is at about a 1:10 ratio, if we simply predict that all the comments given are nondelta-winning comments, then our overall accuracy would be over 90%. However, this does not
accurately represent our data and also does not assist us in finding the most prominent persuasive linguistics.

Some of the main strategies for fixing this are random down-sampling the majority class, random up-sampling the minority class, and Synthetic Minority Oversampling known as SMOTE. Down-sampling is when you grab only a small portion of the total dataset. Naturally, since it would take an incredibly large amount of time to train our algorithms to use all 120,000 nondelta-winning comments, we want to minimize the amount the comments we use. This also allows us to get somewhat closer to the number of delta-winning comments that we have. The other two techniques are up-sampling, or basically increasing the number of delta-winning comments that we have by duplicating them into our dataset. By duplicating them, we can match and equal the number of different comments to remove the bias towards the larger one. The first form of up-sampling is random up-sampling. As the name suggests, it randomly selects a comment to be duplicated. The SMOTE, see (Chawla, Bowyer, Hall, & Kegelmeyer, 2002), however, combines comments of a certain type to generate new ones. This technique usually outperforms random up-sampling and helps to stabilize the accuracy of our predictions. After testing down-sampling, up-sampling, and SMOTE, we found that SMOTE was the best in getting us matching accuracies for nondelta-winning comments and delta-winning comments.
3 Feature Engineering

Once our dataset has been properly evened, we can start feature engineering. Feature engineering is generating data and labels for our predictions based on a raw dataset. These features are the inputs for our predictive algorithms. They are what we are attempting to analyze to determine importance from. These are essentially our persuasive linguistics to help us better understand persuasion.

We separated out features into two groups, one was features based on already known predictive linguistics and the other was using Natural Language Processing techniques. For the predictive linguistics our features are made up of frequency counts and categorizing word choice. As we notice, persuasive linguistics already has a relationship towards when a comment receives a delta, see (Tan, Niculae, Danescu-Niculescu-Mizil, & Lee, 2016). However, we will be going beyond and determine whether we can predict if it receives a delta based on those relationships.

3.1 Features based on NLP techniques

Natural Language Processing is the set of techniques used for computers to turn human natural language into computerized data, see (Bird, 2009). Some of these techniques are called Text Normalization or Word Normalization. These are used to simplify text for a machine learning model to more easily process the word. Most words that we know contain roots and are derivations of other words. These Natural Language Processing techniques can grab the words and turn them back into some form of their original root.
The two main techniques for this are called Stemming and Lemmatization. Stemming typically works by removing words from either the beginning or end of that specific word, removing the prefixes or suffixes of a word. It usually does this by searching through the list of words for common prefixes and suffixes, and then removing them from the original word. This technique doesn’t always work, as some words lose their definition after the removal of a prefix of suffix, however, even past the limitation, it is still a very good way to combine words. Lemmatization attempts a more proper removal of prefixes and suffixes from a word by only removing prefixes and suffixes that do not change the underlying definition of the word. Even though this technique is significantly better, it is definitely still flawed. If, for example, two words have the same spelling but different definitions, this technique has the potential to incorrectly define the word and in turn remove the wrong prefixes and suffixes, changing the word to the definition of the other.

Another common technique that is used is removing stop words. This is another form of pre-processing that is run before any of the prediction models can be used. Removing stop words is a form of filtering through text data in order to remove useless or meaningless words within the text. These meaningless words are referred to as stop words. Stop words are words that are overly used, that entirely lack meaning or do not add any meaning to text. These words typically take up space and waste processing power when running through the machine learning model. Removing these words is a simple process. Using a complete list of stop words, you go through your data of text and delete a word whenever it is a stop word.
Tokenization is the technique of grabbing a chunk of text and separating it into several pieces called tokens. In the process of tokenization, certain forms of punctuation like periods or apostrophes attached to words may be removed. Even though tokens can also be referred to generally as words, these words are a collection of characters which will be then manipulated using other natural language processing techniques.

Once they have been tokenized, we then use a method called vectorization on the tokenized text to generate features. Since we are using models that require numerical values to run our predictions, the features, or inputs that we give our models are always numerical. For this reason, we need to convert our tokens of strings into numerical values. Vectorization is the name of the process that we use to convert those strings into values.

The first type of vectorizer is called Count Vectorizer. Count Vectorizer gives us a simple way to keep track of which words or tokens have been used in each input given to train. This vectorizer goes through all the input and keeps track of the count of all the words and then goes through all the inputs again and applies the counts for every possible token as features to those inputs. Even though word counts are useful and simple, there will always be issues with this, for example words like “this” or “that” will appear several times and their counts won’t necessarily be useful.
TF-IDF Vectorizer, on the other hand, keeps track of term frequency. This is currently the most popular method and stands for “Term Frequency – Inverse Document”. This will allow us to save groups of words that tend to appear often rather than only individual words. TF-IDF becomes a much more efficient vectorizer in comparison to Count Vectorizer.

3.2 Features based on Persuasive Linguistics

Other forms of features can come from persuasive linguistics or understanding the style of the writer. Since our research is more specifically trying to predict how persuasive people can be through text, we can grab persuasive linguistics, or factors that we already know can influence people’s opinions and use those as categories to find out how many of these techniques the input text is using. Some of these techniques include the amount of words you use, the punctuation you use, whether you support the argument with evidence, or even the intent behind word choice.

The basic form of language that we first consider is punctuation. We analyze the text input to identify the different variety of punctuation used as well as the count of each form of punctuation in the text. Once we’re done counting the punctuation, we begin to analyze the lengths of the words within the text. We keep track of lexical diversity, lexical density, and total word count in order to understand how diverse the language of the individual is. The diversity of a person’s language may be a form of showing their intelligence in persuasion. Other examples include character counts and character
diversity, which keeps track of the total amount of characters in the text, and how often each character in the text was used. We also keep track of reply counts, or how often the original poster replies back to you.

Besides counting the words and characters, we can also put each word used in specific categories and keep track of the amount of times a word from a specific linguistic category appears. Some of the categories of words include arousal, dominance, valence, examples, hedges, and self-references. Arousal refers to the intensity of a message or word, ranging from “calm” words to words that “excite”. Implies arousal of emotion, or essentially words that tend to tug on a way a person feels. Dominance represents the strength of a person’s style or assertiveness. Valence is referring to the how pleasant or unpleasant something particularly is. Examples are keywords that imply someone is showing a condition that falls into a pattern or idea (for example, for instance, i.e., e.g.). Hedges are words that allows users to signify how certain or uncertain they are of a specific idea. Self-references are essentially the different ways a person can refer to one’s self.
4 Results

4.1 Models

From our data, in order to predict it properly, we must have several models that we’re willing to try. Below are some of the models that we used.

- Decision Tree – see (Quinlan, 1986)
- Random Forest – see (Breiman, 2001)
- Naive Bayes – see (Xu, 2016)
- Gradient Boosting – see (Friedman, 2002)
- Ada Boost – see (Schapire, 1999)
- Logistic Regression – see (McCullagh & Nelder, 1989)

Once we have created our models and classifiers, we must then determine which few have the highest accuracy. When we initially train our models, select data is given to the models. To predict the accuracy of the model, we give the model a set of inputs, and we compare the prediction values with the actual results. In order to generate our training data, we have to first split our data into train and test sets. Using scikit-learn’s split function, we can both split and randomize our data at the same time. We chose to split train and test into 70 percent and 30 percent. If we were to train the models with the same data that we’re testing the accuracy with, the model would be biased since it would have already seen the true output instead of creating a prediction.
4.2 Accuracy Score/Confusion Matrix

In order to interpret which models are the best ones to use, we use the accuracy score of each one and we compare to see which one is more accurate. The ones with higher accuracy tend to be the ones that will end up being better to use.

Once we determine which classifiers are most accurate, we can then further analyze their accuracies using a Confusion Matrix. This allows us to analyze and compare the accuracies for each delta and nondelta. The confusion matrix gives us the percent of deltas and nodeltas that were correctly and incorrectly predicted. The y-axis contains the true result or what the label was, and the x-axis contains what the model predicted.
Once we have good results with our classification models, we can use some of them to determine feature importance. Feature importance is essentially determining which features have the largest impact on whether a comment will receive a Delta. Looking at the Random Forest Classifier, which gives us the highest accuracy score, we can see the weights or importance for each feature. Below is a chart of the top 10 features.
Besides just importance of the feature, we can determine what percent impact it has on determining the prediction. By averaging the features for each category by delta and nondelta, we can identify the average difference between the two classifications. Below you can see the percent difference of nondelta to delta. The delta column is the average value of that feature for all the deltas, and the nondelta column is the average value of that feature for all the nondeltas. The difference percent is the percent change from the nondelta to the delta. With this, we can see the differences between the two.
<table>
<thead>
<tr>
<th>Feature</th>
<th>delta</th>
<th>nondelta</th>
<th>Difference %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bold</td>
<td>0.27</td>
<td>0.10</td>
<td>166</td>
</tr>
<tr>
<td>Enumeration</td>
<td>0.06</td>
<td>0.02</td>
<td>158</td>
</tr>
<tr>
<td>Character Count</td>
<td>1165.76</td>
<td>496.59</td>
<td>135</td>
</tr>
<tr>
<td>Extremity Count</td>
<td>2.64</td>
<td>1.23</td>
<td>115</td>
</tr>
<tr>
<td>Exclamation Marks</td>
<td>0.13</td>
<td>0.06</td>
<td>110</td>
</tr>
<tr>
<td>Certainty Count</td>
<td>1.68</td>
<td>0.87</td>
<td>93</td>
</tr>
<tr>
<td>Quotes</td>
<td>0.64</td>
<td>0.43</td>
<td>47</td>
</tr>
<tr>
<td>Question Marks</td>
<td>1.03</td>
<td>0.71</td>
<td>46</td>
</tr>
<tr>
<td>Links</td>
<td>0.27</td>
<td>0.21</td>
<td>28</td>
</tr>
<tr>
<td>Average Sentence length</td>
<td>16.33</td>
<td>13.71</td>
<td>19</td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td>72.19</td>
<td>81.64</td>
<td>-12</td>
</tr>
<tr>
<td>Reply Count</td>
<td>0.07</td>
<td>0.79</td>
<td>-91</td>
</tr>
</tbody>
</table>

**Figure 6: Table of classification feature differences**

The graph below shows a two-dimensional model of the top two features. This shows a general trend between delta-winning comments and no-delta winning comments. We see that generally, that deltas have lower reply count.
5. Text Generation

Once we were able to predict with 70% accuracy, we used text generation techniques in natural language processing to create replies to posts based on previous delta-winning comments. To do this we have a pre-trained model called gpt-2 and fine-tuned it to generate the pattern we’re looking for. Gpt-2, see (Radford, et al., 2019), is a model that learned basic reading comprehension abilities by training it with over 8 million web pages from the internet. The idea is relatively simple, based on the previous words that is has seen thus far, and the current word it sees, what is the most likely next word to appear? By generating random text and classifying it based as a reply, we can see how effective the learning algorithm or model really is. We are also able to visually represent weakness in our algorithms by reading and personally evaluating how well a generated comment can change a person’s opinion. For example, a generated comment might repeat
a lot of popular words, add lists and examples, but not make sense in terms of comprehension of the topic.

In order to train this model, we had to give labels for every submission and comment that we hyper train it with. Below is the submission that was given as input to the model.

**CMV submission:**

*Seeing a person dressed as a woman that is clearly a man will never not be weird.*

*I know gender dysphoria is real and transgender people exist. I know that the world would be better off accepting that fact.*

*But in the push to accept transgender people in society, there seems to be this awkward game of pretend that is expected to be played.*

*It is one thing to simply adjust the pronouns you use to make them comfortable, but it is another thing entirely to be expected to convince yourself that they are the gender in which they identify. If you meet a trans woman that is clearly trans (most notably ones that used to be men with clear male facial features, endomorph or even receding hairlines) you are expected to actually believe she is a woman on a biological level. There are some very lucky biological men that can transition and be incredibly convincing, but most aren't, and if you aren't one of those lucky people, you will never be seen as an actual woman no matter how much pretend we are expected to play. We can all act like they are women, but the brain already sorts them out as "men dressed as women" way before any conscious decision is made, and it will always be weird and awkward seeing these individuals no matter how nice and accepting we are about it.*
Below are the three different comments outputted by the trained model.

Output #1

I believe you are missing the point of gender. The notion of gender is a social construct. It is not something you can change. It is a construct that can and should be changed. In my opinion, gender is not a construct you can actually change. It is not something that people are forced to "change". It is a set of behaviors that people can and should mimic.

The issue with this is that you are saying that being forced to be a woman is like being forced to be a man. I am not talking about "being forced to be a man is wrong" or "being forced to be a man is wrong." I am not saying it is okay to be forced to be a man in public, but I am saying it is ok to be forced to be a woman in private. Just think about how uncomfortable it is for you to be forced to be a man in public, and how uncomfortable it will be for you to be forced to be a woman in the same manner. I am willing to bet you that many trans people will not be forced to be a woman, but many will. If you are truly bothered by this, you need to consider what the purpose of the feminine would be in relation to the masculine universal experience, and one that many trans people are forced to face. I am not saying you should not be forced to be a woman, but I am saying you should be forced to be a man. I am not talking here about "being forced to be a man is wrong" or "being forced to be a man is wrong." I am not saying it is okay to be forced to be a man in public, but I am saying it is ok to be forced to be a woman in private. Just think about how uncomfortable it is for you to be forced to be a man in public, and how uncomfortable it will be for you to be forced to be a woman in the same manner. I am willing to bet you that many trans people will not be forced to be a woman, but many will. If you are truly bothered by this, you need to consider what the purpose of the feminine would be in relation to the masculine.
I would go so far as to say that the transgender community has historically, and purposefully, been underfunded, underfunded, and underfunded.

It's actually been a fairly decent start to my post, but at a really low level. I believe it was around $50M for the whole thing. At that point, I was basically looking for a buzzword, and a typo. I think it was around $20M for the first two movies. They made a deal with the studios to make a second Terminator movie, and they were basically giving us a fourth movie. When that came around, the budgets were basically the same, except for a few minor bits.

I think that's why I believe that the transgender community has had such a negative influence on the rest of Hollywood, and the rest of the entertainment industry.

I think that it's been an underfunded community, and I think that it's been underfunded for a long time, but I don't think it was the first time that the LGBT community has been underfunded, either. I think you're missing the point. I think the money was spent on the very first two movies, and that the people that made them, the people that made them financially, the people that made them financially, weren't really paying their bills. That's something that are even more relevant now, because of the influence the LGBT community has had on the rest of the entertainment industry, because you didn't have a chance to really have a say in it.

I think that it's been years of intense negotiation, both within the LGBT community and within the industry itself, and it was basically this:

1. How much money should we make to help and defend the community?
2. What do you think of the current high standards of the industry?

I think that your question is a little fuzzy, and you kind of need to narrow it down a bit. You're really only asking this in the
sense that you're asking it in the sense that you're asking it in the sense that you're asking it in the sense that you're asking it. I think that if you're serious about fighting this, you should talk to some smart people. You need to be willing to negotiate, but you need to be willing to talk to people who are willing to pay for it, and they need to be willing to pay for it. If you're not willing to pay, you're not going to get a decent representation. If you're willing to pay, you're not going to get a good one.

Ultimately, you need to make sure that you're making a good enough case to convince people. If you're not in good company, you should be ashamed of yourself. It's really not a matter of whether or not you're in good company, but who is in good company?

EDIT: Sorry for the lengthy post, but I couldn't find a good post with all of the information I needed. I'll continue to post as I get more and more errors.

Output #3

I'm aware that this may be a little off topic, but I'm not going to lie about my attitude towards this. I'm *not* surprised by it. I've been taught that trans people are basically nothing more than caricatures of men, and thus women. I'm not even sure that trans people can be expected to be feminine in any way, but I assume that's the reason behind my *not* taking gender dysphoria personally.

If I were a cis man, I'd be pissed. I'm not saying that I'm against trans people being able to transition, I'm not arguing. But I'm not saying that *men dressed as women*.

I'm simply saying that my view is *completely* wrong.
All that said, I do think that there is room for a bit of debate on this. I think that *trans* people should be treated equally, though I do think that trans people can be expected to wear the gender-shifting clothing and hairstyles that they choose, regardless of gender. And I think that trans people can be treated fairly, though I am open to that as well.

But this isn't how it works. I've met a lot of trans people, and I've been treated pretty badly. I'm not saying that I'm feeling bad about it, just that I don't think that it's entirely fair. I think that cis people who complain about being treated like shit are just being transphobic, and that's totally unfair.
6 Conclusion

Overall, persuasion is one of those topics that has several factors that need to be considered. In this research, we attempted to classify successful persuasion was accurate to a minimum of 70%. This implies that computers, with the technology we have now, are somewhat able to predict whether a comment deserves a delta but is not a complete tell.

One of goals from this research was to determine the key features in whether a comment receives a delta. From looking at our results relating to our features, we can see that reply count had the highest impact in our random forest classifier. However, after looking at our classification differences, we can see that reply count has a very small number for deltas in contrast to nodeltas. This is something that would have to be further researched to try to identify the cases as to why people who reply more often get rejected and not convinced.

After looking at the number of replies, the next important category of features become our count values. Since the average values for lexical diversity and average sentence count are very close for deltas and nodeltas, relative to the other features, it is hard to determine the impact that those features most likely have on our classifiers. Character count has a large difference between the classifications. We notice from our table of classification feature differences that there are 135% more characters on average for deltas, compared to nodeltas. This implies that on average, delta-earning comments have
more than double the amount of text. There could be many reasons as to why this is the case, such as they use more examples, have larger explanations, use more complete sentences, and other possibilities. To fully determine the reason behind this, it would have to be developed in more research.

To see how human-readable the comments have to be in order to receive a delta from a classification model, we used text-generation techniques built on previous comments to see what comments could receive delta and how accurate this would be to our human eyes. After observing the text generation, we see that it does include things such as examples, lists, references, and are generally long. However, after reading the actual generated comment, we notice that the comprehension level of the model is low, and the sentences lack meaning. They do not target the question at hand very accurately, but instead use features from our classifiers seemly at random. Another idea to further this research would be to build a model with reading comprehension abilities, especially to determine if the “question” has been answered.

The next step in this research would be to generate a much larger list of features using feature engineering techniques, and to build a neural network as a model to determine whether a comment should receive a delta.
References


