Sentiment Analysis of Haribo Gummy Bears

ENSF 612

University of Calgary

December 15, 2022

Chansuck (Shawn) Lee – 30169231 - chansuck.lee@ucalgary.ca

Nicolas Hirschfeld – 30172066 - nicolas.hirschel1@ucalgary.ca

Tania So – 10046662 - tso@ucalgary.ca

Individual Contributions:

Name	Data Collection	Coding	Writeup
Tania	Raw data collection and Labelling of data rows 1-333	Code	Introduction, Results, Discussion (⅓)
Nic	Labelling of data rows 334-666	Code	Conclusion, Results, Discussion (⅓)
Shawn	Labelling of data rows 667-999	Code	Abstract, Results, Discussion (⅓)

Required Links:

Link to Dataset: https://www.amazon.com/Haribo-Gummi-Candy-Goldbears-Pound/dp/B000EVOSE4/ref=cm_cr_srp_d_product_top?ie=UTF8&th=1

Link to Model Implementation: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/929313280972626/2238895283214799/8788303842077579/latest.html

Table of Contents

1 – Abstract ***1***
- Context: 1
- Objective: 1
- Method: 1
- Results: 1
- Conclusion: 2

2 – Introduction ***3***

3 – Results ***4***

a) How was the new data labelled/collected? **4**
- Approach: 4
- Results: 4

c) How was the data preprocessed? **6**
- Approach: 6
- Results: 6

e) How does the performance of the models change based on the choice of hyper parameters? **8**
- Approach: 8
- Results: 9

f) How are the misclassifications of the best performing model distributed? **10**
- Approach: 10
- Results: 11
- Analysis of misclassified records: 17

4 – Discussion ***21***

5 – Conclusion ***23***

6 – References ***24***

1 – Abstract

Context:

An Amazon product's reviews and ratings are an important tool for consumers when making a decision of whether or not to buy a product. Reviews/comments can be especially useful for consumers to read about a specific positive/negative attribute about a product. Oftentimes, there are conflicting sentiments between the review and the rating of a product. For the consumer, this conflict may impede their decision on purchasing the product.

Objective:

In this paper, we study the process of using Sentiment Analysis to predict the sentiment of a specific Amazon product (Haribo Gummy Bears) using just the customer reviews.

Method:

Using 1000 manually labelled data points, we trained and compared a combination of 2 Feature Extraction methods and 3 Machine Learning models to predict and label reviews as irrelevant, negative, neutral, or positive. Bag-of-words and TF-IDF were used for the feature extraction methods while Naive Bayes, Logistic Regression, and Random Forests were used as the ML models.

Results:

We found that using the combination of Bag-of-words and Naive Bayes produced the highest accuracy of labelling (58.64%). The range of accuracy of all tested combinations of feature extraction and model was between 40-60% [Table 1].

Table 1: Final accuracy scores and corresponding models

Feature and Model	Accuracy
Bag of Words and Logistic Regression	52.79%
Bag of Words and Random Forests	50%
Bag of Words and Naive Bayes	58.64%
TF-IDF and Logistic Regression	51.55%
TF-IDF and Random Forests	40.12%
TF-IDF and Naive Bayes	56.41%

The tuning of hyperparameters (smoothing) for Naive Bayes proved to have no effect on the accuracy of labelling; therefore, the default smoothing value of 1.0 was used for the model. Comparing the actual vs predicted label ratings, we observed that the ML model trended to predict a proportionally larger amount of 1s and 3s (negative and positive) [Figure 1].

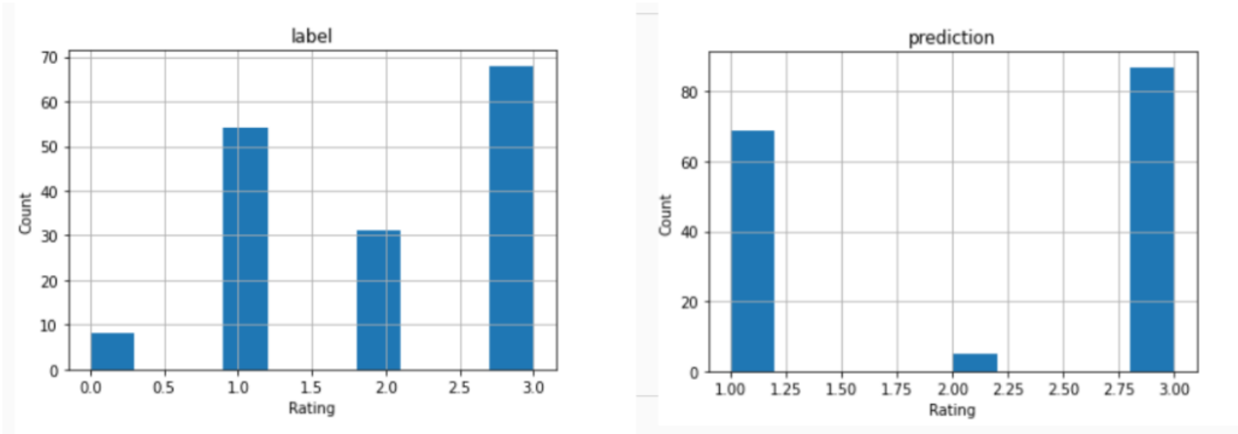

Figure 1: Initial data labels and final predictions

Conclusion:

With more test labels and further analysis/implementation of different feature extraction methods such as n-grams, sentiment analysis of Amazon product comments using Machine Learning can prove to be very useful in the prediction of a consumer's sentiment/rating by just looking at their review.

2 – Introduction

E-commerce has changed the way consumers are able to obtain products. We have access to global supply chains and are confronted with multiple options when searching for a product. It can be overwhelming to decide on one brand and flavour of gummy bears when so many choices are available with one click of a mouse. It is now common practice for businesses to provide a rating and review with a product to help consumers with decisions. A seemingly trivial purchase such as gummy bears can be selected with the highest rating out of 5 stars, and we can free our mental capacity for more important matters. There are thousands of contributions towards the 4.6 star rating that provides confidence towards the purchase. However, upon reading the reviews, the consumer may find that the trivial purchase of gummy bears is not a harmless one.

The project explores whether sentiment analysis on Amazon reviews can provide a useful tool to assist consumers in purchasing decisions. At times, a 5 star rating can be accompanied by a negative review and vice-versa, where the ratings and reviews do not align and provide an accurate sentiment on the product. Sentiment analysis would provide an alternative metric by summarizing whether the product is negative, neutral, or positive.

Amazon Haribo Goldbears is listed as a Best Sellers with a 4.6 out of 5 stars rating (32,670 contributions). Through the Amazon Reviews Scraper API, we collected approximately 8000 data items of consumer ratings and reviews. Of this new data, we labelled 1000 points as irrelevant, negative, neutral, and positive (encoded as numeric values 0, 1, 2, 3 respectively). Prior to feeding into our pipelines, the data was preprocessed using standard natural preprocessing methods such as tokenizing, removing stopwords, punctuation and noise. By exploring two feature extraction methods and three classification methods, six different pipelines were used – where the highest performing model was chosen for hyperparameter tuning. The performance of the model did not change based on the choice of hyperparameters. Details of the complete process will be discussed in the following sections.

3 – Results

a) How was the new data labelled/collected?

Approach:

The data was collected through an API (Amazon Reviews Scraper) that obtained 7670 data items in excel, csv, and json formats. The relevant data included rating, title of the review, and description of the review. To select for 1000 points to label, the data was filtered and approximately 333 of 1-Star ratings, 333 of 5-star ratings, and 333 of moderate 2 to 4-Star ratings were selected. This is to increase the quality of our labels by exposing the ML models to the full range of different sentiments to learn from. The initial intention was to label the data with a sentiment as a value between 1 to 5, with a lower score indicating negative sentiments and a higher score indicating positive sentiments. For simplicity, we decided to label reviews as irrelevant, negative, neutral, and positive instead, encoded with values of 0, 1, 2, and 3 respectively. This would also contribute to the quality of our labels by reducing ambiguous intermediate points (2s and 4s).

The labelling method is as follows: An initial labelling guideline was developed as a group to define the criteria for each score (0, 1, 2, or 3). The allocated 1000 data was randomised, and 30 data points were selected where each member labelled the same 30 items based on title and description and hiding the 5 Star rating to prevent bias. Solutions were compared and disagreements resolved to obtain a group consensus and achieve 100% agreement. The initial labelling guideline was revised, and the data was split into thirds to be labelled by each member. To ensure quality labelling, members followed the revised guideline and labelling was done in the presence of the group. This allowed for group consensus on labels if any one member felt uncertain on a label and required consultation.

Results:

Table 2: Revised labelling guideline.

Sentiment	**Label**	**Criteria**	**Example**
irrelevant	0	No mention of the product, incomplete/nonsensical statements	“Actually, I had them sent to my grandson. He is in the Air Force stationed near Shaw AFB, SC”
negative	1	Obviously negative, or negative outweighs positive	“Flavor, texture is really bad. Has been great in the past, but just crap now”
neutral	2	Mentions product, but neither positive or negative, or both positive and negative where it balances out, eg. intention of getting sick but didn’t, good gummies bad packaging	“I love these but this bag was a stuck together and a bit hard. Not as fresh as usual”

positive	3	Obviously positive, or positive outweighs negative	"These are very good! I recommend them to anyone"

Reviewer 1 Score	Reviewer 2 Score	Reviewer 3 Score	Agreeance	reviewDescription
1	1	1	1	These are the worse gummy bears I have ever eaten. They have no flavor or taste. Won't buy again
3	3	3	1	Just what I wanted.
1	1	1	1	TRY TO EAT IT IN ONE SITTING, AND YOU WILL BE SITTING ON THE TOILET FYI!!!!
3	3	3	1	Fresh and an all time favorite!!
1	1	1	1	I was disappointed in the taste of these bears. Something was just "off". Threw out the bag. Also missing the clear
1	1	1	1	Well.... the Haribo Gummi Candy Gold-Bears, 5-Pound Bag was kind of a joke to get. First they do not have too m
3	3	3	1	I just love the fruitty flavor and the texture
2	0	2	0	What can I say, 5-pounds of bears
3	3	3	1	These are the good ones! Not the ones that REALLY taste fake. (Gummy fans will know what I mean)
2	2	2	1	i read the previous reviews about the Brazilian Haribo gummies and I thought.."well, how much different could t
3	3	3	1	Easy transaction, no problems at all !!!!!!!!
3	3	3	1	you won't be disappointed in this. the seller shipped it packaged well, it arrived earlier than expected and is exac
1	1	1	1	Ordered these numerous times but this bag is from Brazil. They're smaller and the taste is not good. Ashame Harit
3	3	3	1	yum, yum, yum, and more yum, yum.
3	3	3	1	Um. 5 pounds of Haribo. Any person who is serious about there gummy bears knows why its Haribo or nothing.

Figure 2: Sample of initial 30 data points labelled by each member and the agreeance.

Table 1 above defines the revised labelling guideline with examples. It was revised following an initial labelling process where 30 data points were labelled by each member and the agreement was compared (Figure 1). In the agreeance column in Figure 1, a score of 1 indicates that all members supplied the same label, whereas a score of 0 indicated that there was a discrepancy with at least one member. Of the 30 labels, 20 were in agreement giving an initial agreeance of 66%. The 10 labels with disagreements were highlighted and discussed amongst the group to arrive at a consensus (and revise the labelling guideline) until 100% agreeance was achieved. For example, the highlighted disagreement above: two reviewers scored the description as neutral, 2, while one reviewer scored the description as irrelevant, 0. The group came to a consensus that the label should be neutral, 2, as the review explicitly mentions the product, making it relevant, but the comment was neither positive nor negative.

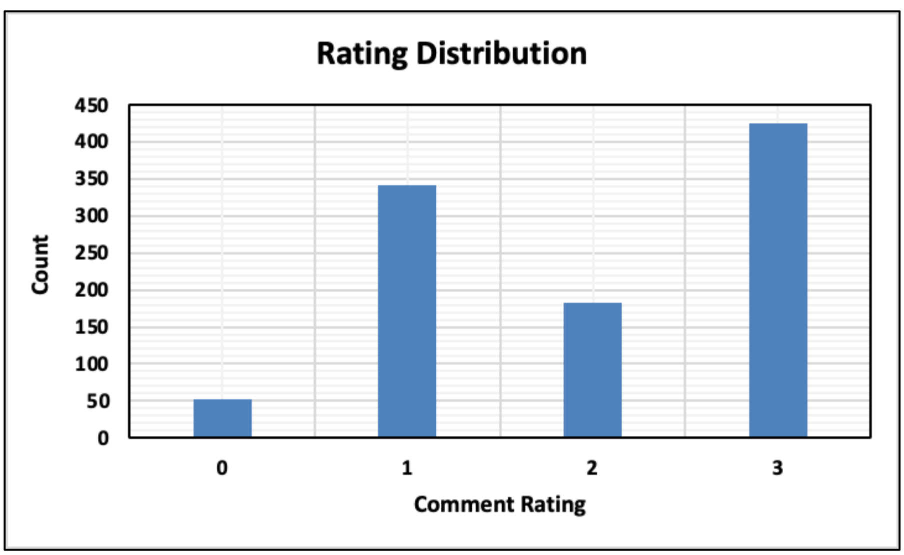

Figure 3: Distribution of labels.

Once the group was confident in the revised labelling guideline, the remaining data was split into thirds for labelling. Figure 2 above shows the distribution of labels: 52 items were labelled as 0 (irrelevant), 341 items were labelled as 1 (negative), 182 items were labelled as 2 (neutral), and 425 items were labelled as 3 (positive).

To improve on the quality of our labels, the data was initially filtered to include an equal amount of negative (1 Star), neutral (2-4 Stars), and positive (5 Star) ratings to label. This is to expose the ML models to the full range of sentiments. Also, labels were reduced from an initial 6 categories (0-5) to 4 categories (0, 1, 2, 3) to reduce ambiguity and increase consensus. We also labelled the 1000 data points in the presence of the group, to provide feedback when needed and thus increasing agreements.

c) How was the data preprocessed?

<u>Approach:</u>

Preprocessing steps include combining the two relevant columns 'reviewTitle' and 'reviewDescription', followed by removing stop words via Pyspark's StopWordsRemover, and noise reduction, where non-alphabetic characters (including punctuation) and words less than 3 characters are removed. Stop words and noise serve to satisfy grammatical rules and sentence structure but do not contribute meaningfully to sentiments.

Additional preprocessing steps include spell checking, and lemmatization. In natural language processing (NLP), lemmatization is used to convert groups of words to their base root-form. Compared to stemming, which cuts the word to its base (studies = studi), lemmatization uses a dictionary-based approach (studies = study) that provides more accuracy when the meaning and sentiment of the word is important for analysis. Lemmatization is a valuable preprocessing step for feature extraction and creating a term document for the corpus of data.

<u>Results:</u>

Figures 3 to 7 below demonstrate the preprocessing steps and the intermediate results. First the relevant columns reviewDescription and reviewTitle are combined (Figure 3). Then stop words are removed using PySpark StopWordsRemover (Figure 4) where the words are tokenized for analysis. An indicator of its effect is the removal of the words "I" and "the". Then noise is filtered out (Figure 5) where we see "m" and "re" are removed. This is followed by spell checking (Figure 6) where "cubies" is replaced by "cubes". Finally lemmatization to produce our preprocessed text (Figure 7), where we see pluralized words are reduced, for example "bears" to "bear".

reviewDescription	reviewTitle	Combined
I recently ordered a 5 pound bag of Goldbears, which came in the original gold bag and says "Made in Turkey." Those were great and I truly enjoyed them...5 stars.This time, I ordered the same product, but received a clear bag of "Goldbears" that says "Made in Brazil." I am very disappointed, and almost feel as if I've been scammed. These bears are much softer in texture than the gold bag, taste is disappointing combared to the gold bag, and are overall inferior to the "Made in Turkey" gold bag p...	No guarantee which version of product or source. Feel ripped off...	I recently ordered a 5 pound bag of Goldbears, which came in the original gold bag and says "Made in Turkey." Those were great and I truly enjoyed them...5 stars.This time, I ordered the same product, but received a clear bag of "Goldbears" that says "Made in Brazil." I am very disappointed, and almost feel as if I've been scammed. These bears are much softer in texture than the gold bag, taste is disappointing combared to the gold bag, and are overall inferior to the "Made in Turkey" gold bag p...
Used this big bag to fill a Mason jar for a baby shower "count the cubies" super cute and colorful.	Love gummy bears	Used this big bag to fill a Mason jar for a baby shower "count the cubies" super cute and colorful. Love gummy bears
If you get a Haribo from grocery store it'll be made in Germany and delicious. These are made in Brazil and don't taste the same at all. Packaging is just a transparent bag, no colorful branding.	It's not same with the grocery store item	If you get a Haribo from grocery store it'll be made in Germany and delicious. These are made in Brazil and don't taste the same at all. Packaging is just a transparent bag, no colorful branding. It's not same with the grocery store item
I'm not the only one who likes these golden bears	Four Stars	I'm not the only one who likes these golden bears Four Stars

Figure 4: Combining reviewDescription and reviewTitle columns

No Stops

["recently", "ordered", "5", "pound", "bag", "Goldbears", "came", "original", "gold", "bag", "says", "Made", "Turkey", "great", "truly", "enjoyed", "5", "stars", "time", "ordered", "product", "received", "clear", "bag", "Goldbears", "says", "Made", "Brazil", "disappointed", "almost", "feel", "ve", "scammed", "bears", "much", "softer", "texture", "gold", "bag", "taste", "disappointing", "combared", "gold", "bag", "overall", "inferior", "Made", "Turkey", "gold", "bag", "product", "won", "re", "order", "anytime", "soon", "gamble", "product", "receive", "guarantee", "version", "product", "source", "Feel", "ripped", ""]

["Used", "big", "bag", "fill", "Mason", "jar", "baby", "shower", "count", "cubies", "super", "cute", "colorful", "Love", "gummy", "bears"]

["get", "Haribo", "grocery", "store", "ll", "made", "Germany", "delicious", "made", "Brazil", "taste", "Packaging", "transparent", "bag", "colorful", "branding", "grocery", "store", "item"]

["m", "one", "likes", "golden", "bears", "Four", "Stars"]

Figure 5: Removing stop words from combined text

No Noise

recently ordered pound bag goldbears came original gold bag says made turkey great truly enjoyed stars time ordered product received clear bag goldbears says made brazil disappointed almost feel scammed bears much softer texture gold bag taste disappointing combared gold bag overall inferior made turkey gold bag product won order anytime soon gamble product receive guarantee version product source feel ripped

used big bag fill mason jar baby shower count cubies super cute colorful love gummy bears

get haribo grocery store made germany delicious made brazil taste packaging transparent bag colorful branding grocery store item

one likes golden bears four stars

Figure 6: Removing noise

Spell Check
[recently, ordered, pound, bag, goldberg, came, original, gold, bag, says, made, turkey, great, truly, enjoyed, stars, time, ordered, product, received, clear, bag, goldberg, says, made, brazil, disappointed, almost, feel, scammed, bears, much, softer, texture, gold, bag, taste, disappointing, compared, gold, bag, overall, inferior, made, turkey, gold, bag, product, won, order, anytime, soon, gamble, product, receive, guarantee, version, product, source, feel, ripped]
[used, big, bag, fill, mason, jar, baby, shower, count, cubes, super, cute, colorful, love, gummy, bears]
[get, harbor, grocery, store, made, germany, delicious, made, brazil, taste, packaging, transparent, bag, colorful, branding, grocery, store, item]
[one, likes, golden, bears, four, stars]

Figure 7: Spell checking

Preprocessed
recently ordered pound bag goldberg came original gold bag say made turkey great truly enjoyed star time ordered product received clear bag goldberg say made brazil disappointed almost feel scammed bear much softer texture gold bag taste disappointing compared gold bag overall inferior made turkey gold bag product won order anytime soon gamble product receive guarantee version product source feel ripped
used big bag fill mason jar baby shower count cube super cute colorful love gummy bear
get harbor grocery store made germany delicious made brazil taste packaging transparent bag colorful branding grocery store item
one like golden bear four star

Figure 8: Final lemmatization step

e) How does the performance of the models change based on the choice of hyper parameters?

Approach:

Various hyper parameters that are specific to the classifiers were chosen to be tuned. The logistic regression model used regParam and elasticNetParam for tuning. The random forest model used maxDepth, maxBins and numTrees for tuning. Lastly, the naive bayes model used smoothing for tuning. Refer to the table below to hyper parameters that were tuned for each model and their corresponding default values.

Table 3: Tested models and corresponding hyper parameter tuning

Model	Hyper Parameter Values	Default Hyper Parameter Value
Logistic Regression	regParam [0.0, 0.3, 0.5] elasticNetParam [0.0, 0.1, 0.2]	regParam = 0.0 elasticNetParam = 0.0
Random Forest	maxDepth [2, 5, 10] maxBins [5, 20, 32] numTrees [5, 20, 50]	maxDepth = 5 maxBins = 32 numTrees = 20
Naive Bayes	smoothing [0.0, 0.2, 0.4, 0.6, 0.8, 1.0]	smoothing = 1.0

The regParam for logistic regression is a regularization parameter. Regularization discourages learning a complex or flexible model, which mitigates the risk of overfitting. ElasticNetParam is the elastic net mixing parameter for logistic regression. By changing the elasticNetParam from 0 to 1, the penalty gets shifted from an L1 penalty to an L2 penalty. Essentially, a weight is being assigned to the L1 and L2 penalties through changing the elasticNetParam.

The maxDepth parameter for the random forest classifier specifies the maximum depth that the tree is allowed to grow. Complexity and time increase as the tree increases in depth. Training error will never go down as the depth increases. However, training error may also never go up if the hyperparameter has peaked, overall, increasing the model complexity with no benefit. Therefore, the desired maxDepth is one which will allow an optimised training error without increasing the model complexity significantly. MaxBins parameter represents the number of bins used when approximating continuous features. MaxBins determines how to split on features at each node. The last parameter that was changed during random forest classifier tuning was numTrees. NumTrees represent the number of trees specified in the model. Generally, more trees will result in a more accurate model. However, increasing the number of trees will also result in an increase in computational cost. Therefore, there will be a peak numTrees value for the model because the accuracy improvement will be negligible compared to the increase in computation cost.

The only common hyper parameter to tune for the naive bayes classifier is smoothing. Smoothing mitigates the problem of having a zero-probability result with the naive bayes classifier. Producing a probability of zero is unrealistic and can cause computation challenges. Smoothing will increase the resulting probability just above zero to mitigate these issues.

Results:

As previously mentioned, there are six final cases that were modelled and optimised. Through tuning with Naive Bayes while using bag of words as the method of feature extraction, the highest accuracy score of 58.64% was achieved. The same score was obtained by using bag of words in combination with a random forest classifier.

Table 4: Accuracy scores for each model case at the optimum tuned hyperparameters

	Feature Extraction and Classifier	Accuracy Score	Hyperparameters
1	Bag of Words + Logistic Regression	0.5679012345679012	regParam: 0.3, elasticNetParam: 0.1
2	Bag of Words + Naive Bayes	0.5864197530864198	smoothing: 1.0
3	Bag of Words + Random Forest	0.5864197530864198	maxDepth: 10, maxBins: 5, numTrees: 50
4	TF-IDF + Logistic Regression	0.5679012345679012	regParam: 0.3, elasticNetParam: 0.1
5	TF-IDF + Naive Bayes	0.5617283950617284	smoothing: 1.0
6	TF-IDF + Random Forest	0.41358024691358025	maxDepth: 10, maxBins: 5, numTrees: 5

Table 4 represented above shows the results from the six cases examined along with the optimised hyperparameters that were selected for each best-case scenario found through tuning. It is evident that the only hyperparameter that was maintained at the default value was smoothing for Naive Bayes cases. All other hyperparameters were optimised through tuning. One observation that can be made is the trend for the hyperparameters for the random forest classifier cases. The max depth and max bins remain the same as 10 and 5 respectively even when equipped with different methods of feature extraction. Meanwhile, the optimum number of trees changes from 50 to 5 when using bag of words to TF-IDF. However, the accuracy score when using TF-IDF is generally lower than when using bag of words, especially when a random forest classifier is used. TF-IDF equips a high weight to words that do not appear often throughout the amazon comments. Therefore, “rare” words will have a higher significance on the model than words that are frequently used. Bag of words essentially does the opposite. Bag of words associates high significance to words that are frequently used amongst the data. Rare or uncommon words do not need high significance for sentiment analysis of amazon comments. The common words such as “good” or “bad” are more important, which supports the use of bag of words being an optimum feature extraction method. The results displayed in table 3 correlate to bag of words being a more accurate feature extraction method than TF-IDF.

f) How are the misclassifications of the best performing model distributed?

Approach:

The best performing models were the pipelines containing Bag of Words for feature extraction and Naive Bayes, or Random Forest as the ML model giving a prediction accuracy of 58.64%. The predicted output is obtained for each data point and compared to the target label. Distributions of both the labels and the predictions are also compared. With a 58.64% prediction accuracy on the test data, it is evident that there are many incorrectly labelled comments, even on the best performing models. Some example cases will be analysed below to demonstrate why the model was wrong. Results for the Bag of Words combined with Naive Bayes model will be used for analysis.

Results:

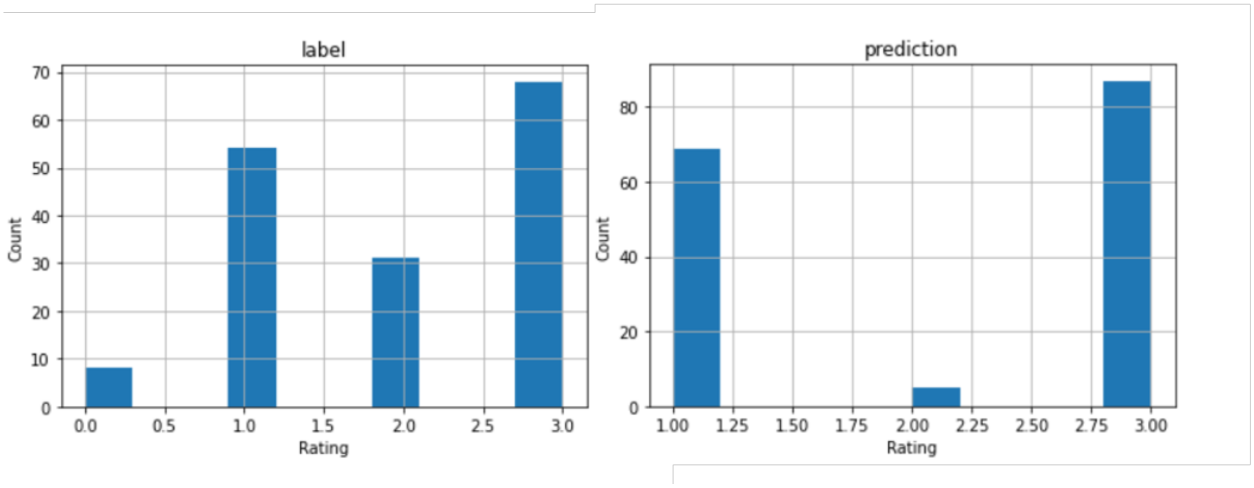

Figure 9: Initial label distribution and final predicted

The results of the model's predictions show that it predicted more 1's and 3's (negatives and positives) and less of the 0's and 2's (irrelevant and neutral) [Figure 8]. Below, we have picked all of the misclassified records and displayed them as a table. There is also an analysis of why each of the records were misclassified.

Num	Preprocessed	Label	prediction	match
1	admit first thing got bag gummy bear look made confirm one made germany actually forgot check product made because wanted harbor gummy bear check review photo taste fine dumped big bowl snack throughout day think fine snacking yes made germany still good	2	1	FALSE
2	although item arrived fairly quickly time placed order disappointed opened bag first bag tore open completely meaning find large container transfer dummy noticed huge chunk dummy mangled stuck together perhaps due humidity poor packing shipping dumped mangled bunch cabbage probably losing one pound dummy disappointed taste flavor better dummy occasionally buy co wonder relatively cheap packaging used dummy compromised freshness flavor love dummy wait order head nearest pharmacy candy store	1	3	FALSE
3	amazon review bit confused many review product mention bowel churning property sugar free gummi bear sugar free normal gummi bear intended human consumption buy want enjoy gummi bear fiercely colon emptying sugar free gummi bear far far expensive explicitly labeled mislead review product make explode induce horrible horrible diarrhea	2	1	FALSE
4	aren fresh expected gummi bear always delicious kid aren picky though kept anyway frutti gummi bear	2	3	FALSE
5	arrived really quick hard tho supposed dummy hard candy	2	3	FALSE

6	ate nickel two found shortly eating bag five star	2	3	FALSE
7	bag got seemed mostly lemon lime gummy bear flavor good expected fine kid grandkids like adult disappointed even kid said decided order update grandkids ask order personally didn care willing buy anyway say much cost said way check around online got first time yes said way pay much worth shocked price jump thing amazon seems excessive taste good expected	1	3	FALSE
8	bag open disappointed gummi bear fresh bag arrived happy	2	1	FALSE
9	bear fresh arrived amazon price way better pay super market think palmar charge buck bag little tax steal thought smell bear first time bought pounder think used many like sensory overload happy every time order harbor bear sure side bag bear repeatable like bag never real problem transfer plastic ziplock great value taste	2	3	FALSE
10	bought investment fid gold bear contain gold whatsoever waste money contain bear listed among ingredient taste like moose maybe little like boar definitely like bear totally false advertising warning contains neither gold bear	0	1	FALSE
11	bought several time usually great last time aren fresh straight bag kind hard still edible doubt certainly fresh probably give one shot fresh time stale won buy anymore sometimes stale	2	3	FALSE
12	box came sealed bag containing gummi open entire box arrived sounded like box nail contemplated hygiene ended giving toddler ate without negative effect pretty sad story though reality imagine going grocery buying bag candy break open front porch many got dirty dirty get lose poor packaging	2	1	FALSE
13	buyer beware noticed reading review received clear bag product different substandard version one harbor bag best gummy bear look	2	1	FALSE
14	didn taste like hairdo sure purchase	2	1	FALSE
15	dummy guess really expecting sugar free variety harbor discontinued line bought gag gift brother got regular sugary dummy well bit disappointed tasted guess	2	1	FALSE
16	fantastic lovable pound bear arrived heavy duty clear plastic bag ingredient nutritional information plain white able admit eating one dropped counter top transferring new sealed plastic container flavor texture may original far concerned introduced visiting parent german summer break vending truck made round thou military housing area every exactly his treat year old brother took nap definitely addictive sixty year passed since remain one favorite original delighted fresh right texture flavor name harbor country origin given single label ingredient nutritional information seller listed amazon lac harbor eligible return either product information description listed harbor brand name including image wherever bear came least arrived safely safely time first order went missing update honestly expected yummy bear last longer one half week whatever ingredient suppressing coughing allergy exceeded cough medicine medication using opposed cough suppressing lozenge definitely effective tract upset negative side effect certainly isn recommendation others simply observation terrific price great flavor texture	3	1	FALSE

17	felt good buying picture showed made india arrived made brazil package knew want one brazil several negative comment false advertising recommend buying local store sure come bought fresh market made germany manufactured place written package	2	1	FALSE
18	first couldn believe sold gummy bear bag need pound gummy bear saw available subscribe save need subscribe gummy bear shipment couple day later needed subscribe came part subscribe save monthly shipment look exactly pictured largest bag gummy bear ever seen opened dumped gallon ziploc bag fit well stayed fresh past month nice able pantry eat couple gummy bear time never eat little bag want uneaten one bad even happens never around long enough find know need pound gummy bear every month hurt get couple time year right subscribe save let choose delivery schedule never without gummy bear subscribe save gummy bear	2	3	FALSE
19	flavorful sticky difficult tell different flavor purchase sticky	2	3	FALSE
20	four star	2	3	FALSE
21	four star	2	3	FALSE
22	god buy spreading diabetes household ordered bag get two day later week handful bear grazing hubby culprit leave bag go work come home baggie lunch video game dinner dinner movie sex teeth brushing teeth brushing handful way shower won end sake hope long healthy marriage last time buy gummy bear large delicious fresh quantity god buy	2	3	FALSE
23	good dummy rather firm considering expiration date year away sure must manufactured within soft cheapness noticeably fresh dummy know great feeling poke package know won wreck teeth allow warm room temperature eating almost feel though benefit zap microwave firmer	1	2	FALSE
24	good thing werent look explosive review sugar free one whew coworker kindly offered delicious gummi bear another worker pregnant worked said really taking gummi bear pregnant lady yes like gummi bear heroin addiction need sustain life shipwreck none le hatched devious plan replenish gummi bear supply pregnant lady happy said bag open wanted first person open bag take stiff bowed pregnancy police worker winning sugar free	0	3	FALSE
25	got gummy bear noticed something wasn right bag weighted pound began cry pound instead really good sorry spell thing wrong bow case want send another one contact email send make sure pound get make sure pound please make cry pound	1	3	FALSE
26	got wanted try flush everything guy luck thing guy irritate stomach caused bloating gave lot gas tasted good finally understood review saying dumb chewy thought exaggerating first case think purchase meh	2	1	FALSE
27	great harbor gummy bear however first time bought product repeatable bag second time least third time clearly state packaging may either familiar retail bag zipper top generic plastic bag product label stuck certainly mind packaging find little absurd need tape bag closed maintain freshness biggest issue packaging	3	1	FALSE

28	gummi bear know scooped large helping put syllable cake sized bowl still half left bag bag need sealed else dry ordered also open order pair expensive headphone whoever packed used gummi bear bag cushioning move saved packaging order july summer candy melted left sticky mess everywhere gummi bear know	2	3	FALSE
29	gummy bear around house many year want reveal make feel old dated wife love grandkids neighbor friend kid love seems like keep enough house friend kid show front porch think seeing smile family friend kid face replaced enjoyable	0	3	FALSE
30	halloween occasion alright buck mean expecting world class candy taste anything make special people already said special balance sweet chew ability anyhow considering buy ton candy buck expectation sadly met read harbor world class gummy bear see status implies higher value expectation exists regardless service halloween time well take care people beware give teeth odd sensitivity harbor world class	1	3	FALSE
31	harbor gummy bear eat one ummm gummy bear	3	1	FALSE
32	harbor make arguably best gummy bear giant bag really convenient use gummy bear house like shot block since contain basically ingredient body need quick energy sugar sodium last time ordered may shipped free amazon prime shipping contacted amazon say control unfortunate anyway get cheaper somewhere else update price reduced change available prime waiting reduction order price increased double	2	3	FALSE
33	huge bag candy foster waistline help keep dentist business really great one problem bag material flimsy easily tear bag open candy fall avoid spillage put bag inside large zipper plastic bag open candy also stored inside sealed bag always fresh gummy bear even bag open huge bag last forever	2	3	FALSE
34	kid like fine wasn overly impressed fresh melted together large clear plastic bag hole tear flavor pretty good pretty evenly distributed exception clear one son favorite weren many found stiff chewy eating small handful ended hurting jaw noticed kid aren eating fast either mostly son us make edible slime friend big hit flavor pretty good start try edible slime kid lot leftover large bag cup gummy bear tablespoon powdered icing sugar tablespoon cornstarch make decent edible slime	2	3	FALSE
35	know buy failure life gummy bear indeed gummy bear	0	1	FALSE
36	know got bad bag good review read toughest gummi bear ever hard chew point mouth begin hurt still four pound remaining couple month since aren enjoyable eat taste isn greatest either unless let set mouth toughness make enjoyable maybe bad bag reason say good weren experience far tough chew	1	3	FALSE
37	last time ordered gummy bear came fresh soft time got hard delivery gummy bear came little hard	2	3	FALSE
38	like remember taste black forest one way better writer opinion childhood favorite prove remembered	1	3	FALSE
39	love bear love golden gummi still maybe couple argh leave golden bear place betwixt heaven hell like plague golden fruit flavored locust hive bear descended upon moment weakness suffered pain glory harbor golden overload	0	3	FALSE
40	love harbor bag definitely stale side little disappointed little stale	2	3	FALSE

41	love harbor gummy bear buy smaller container clearly fresh one sold smaller container dry rubbery instead moist like one buy smaller pack fresh flavor	1	3	FALSE
42	love ordering gummi bear every single time bag come open packaging used fragile guy make adjustment bag fragile broke two three time	2	3	FALSE
43	love product must received older product good first batch ordered enjoy though definitely pay attention ordering next time love product must received older product	3	2	FALSE
44	love thing going hit order button saw went dollar month normally see last month doesn make sense cost	2	3	FALSE
45	made brazil normally come turkey without side side comparison couldnt done seemed smaller flavor lower still quality product possible different ordering new order come brazil shall return brazilian manufacturer	2	1	FALSE
46	maybe shower curtain hobbit maybe edible eat plastic clear taste like rainbow strawberry unicorn fusion berry probably taste like look taste like might surprised sent big piece clear plastic like table cloth size gummy bear ordered	0	1	FALSE
47	mom gold bear berry peach pink white grapefruit fruit salad sour gold bear new sweet sour snake hand best gummy candy ever market real harbor best selling candy also desirable replicate sell online accountability beyond semi possible refund fake harbor candy sold seller difficult packaging replicate especially bag doesn smell taste look right eat received imposter gold bear bag fulfilled harbor bear eaten decade taste different consistency white bear spit green bear family felt sick eating refund cover fake harbor candy come country world anything want test best gummy candy authentic harbor	3	1	FALSE
48	much good four star	2	3	FALSE
49	one received brazil love brazil candy taste texture one used germany austria gold bag better genetics brazil get original	2	1	FALSE
50	ordered bag soft hard like tend gold bag still eat either way soft hard eat	2	1	FALSE
51	picky come gummy candy admit harbor gold bear one type gummy bear eat absolutely love ingesting kind gummy bear may able trick eye soon make tongue mere half second away dry heave journey air ground basically gold bear dummy soul first saw option order bag le bag dog food equal weight quite literally gasped inside welled happiness suddenly little voice inside head telling good true must brand battled thought mere possibility dream coming true overwhelming placed order waited impatiently day day amazon said package arriving saturday think nothing else finally saw ups truck nearing home four house away three two wanted run towards knew wouldn get faster waited suspense eating away entirety ups man stepped truck holding box opened door shouted hello must thought young child trapped inside man body waiting new toy grabbed box ran inside tore open beautiful sitting arm ran kitchen found scissors immediately slashed open bag grabbed handful raised mouth half second passed one second two time soul confirmed true harbor goodness two handful thrown mouth found holy gummy grail doubter know truth	0	1	FALSE

52	probably purchased dozen bag year freshness flavor haven always par overall like candy much patient certainly surprised harbor customer service responded several week concern regarding foreign matter see picture familiar recognize color candy material similar nature candy feel edible least caution unsupervised consumption child customer service response foreign substance	0	1	FALSE
53	product keep long time resealed however built sealer bag good taste	2	3	FALSE
54	product subscription usually arrives time enjoyed husband know pandemic brand dummy substituted harbor quality wish amazon inquired whether wanted le normal product gummi bear harbor	2	3	FALSE
55	rated value flavor higher arrived fresher definitely brand claimed many item amazon genuine also knock offs product sold cheaper wish arrived softer somewhat hard stale	2	1	FALSE
56	received mail package broken gummi bear package sad day able salvage happy besides gummi bear tasted good one still packaging packaging broken	2	1	FALSE
57	shipped marine afghanistan one item requested wish list think well received stop sending summer gummy bear troop	0	3	FALSE
58	shipping method sending via mail successful sat hot metal mailbox couple day texas heat town melted possible bear wet slimy haven opened bag yet hoping get texture back doesn look like work bad shipping method	1	2	FALSE
59	sweet	2	3	FALSE
60	sweet treat tough sugar free one stopped giving kid sugar good	2	3	FALSE
61	taste great believe stale even though come good date date actually printed bag sticker bag educated guess candy old stick date make seem fresh better going sam club costco one thing hate online shopping thing worth buying person taste great believe stale even though	2	1	FALSE
62	taste texture nothing like one get germany outstanding bag say made brazil beat nothing much taste texture nothing like one get germany	2	1	FALSE
63	thesis last four star	2	3	FALSE
64	thought ordering weather cooler package came product stuck together large lump order hot weather	0	2	FALSE
65	took week consume yum gummi	2	3	FALSE
66	warning cause rectum enter turbo mode empty entire life rear try eating moderation try waiting day eating matter try inevitable destroy inside plus side best laxative buy seriously clean entire colon want lose lab minute eat bunch wait result taste great end result	1	3	FALSE

67	well subscription couple year package received indicates made brazil turkey people know big reason buying turkish made bear halal using gelatin sourced beef pork quite maddening sourced channel guarantee turkish made bear subscription cancelled bear usually get inferior gelatin source	1	2	FALSE

Analysis of misclassified records:

1. Correct label: neutral. Prediction: negative. **Reason: The word "Germany" was associated in training data as negative.**
2. Correct label: negative. Prediction: positive. **Reason: Since n-gram was not implemented, "although item arrived fairly quickly" was not read in context as a transitional sentence to a negative sentiment.**
3. Correct label: neutral. Prediction: negative. **Reason: Negative keywords found in the sentence like "horrible", "explode", "diarrhoea". Although the sentence, when read in context, has a mix of negative and positive sentiment, the model could not recognize the duality.**
4. Correct label: neutral. Prediction: positive. **Reason: Lots of spelling errors within the text, leading to a more positive sentiment.**
5. Correct label: neutral. Prediction: Positive. **Reason: Due to not using n-grams, context of the sentence was not recognized leading to a positive sentiment.**
6. Correct label: neutral. Prediction: positive. **Reason: Sentence had words "five" and "star", but model was not able to decipher the negative and neutral context.**
7. Correct label: negative. Prediction: positive. **Reason: There are alot of positive keywords like "good", "expected", "fine" but without using n-grams or context of the sentence, the model could not understand that "shocked price jump" or "excessive taste" was a negative sentiment.**
8. Correct label: neutral. Prediction: Negative. **Reason: Keywords like "disappointed" and "open" had a heavier weighing on the sentiment than "fresh", "happy". It should be neutral because of the even appearance of both positive and negative words.**
9. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear. Keywords like "fresh", "happy" had a heavier positive weight on the sentiment.**
10. Correct label: Irrelevant. Prediction: Negative **Reason: Without context, there are many single negative keywords like "waste" or "warning".**
11. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear. Keywords like "great" and "fresh" had a heavier positive weight.**
12. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative sentiments appear.**
13. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative sentiments appear.**
14. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative sentiments appear.**
15. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative sentiments appear.**
16. Correct label: positive. Prediction: negative. **Reason: Very sarcastic sentence full of positive words but in context it is a negative review.**
17. Correct label: neutral. Prediction: negative. **Reason: Keywords like "brazil" and "india" have a negative sentiment from other reviews. Should be neutral from the context of the sentence though.**
18. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear.**

19. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear.**
20. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear.**
21. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear.**
22. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear.**
23. Correct label: negative. Prediction: neutral. **Reason: Without using n-grams/context, it is difficult for the model to label as negative for this review. Reviewer also used many oxymorons like "soft cheapness" and "fresh dummy".**
24. Correct label: irrelevant. Prediction: positive. **Reason: Without context, the model just reads the single positive keywords such as "good" or "delicious" but when read in context, it is a very irrelevant story.**
25. Correct label: negative. Prediction: positive. **Reason: Rare negative words like "cry" appear in this text. Model is not trained enough to know "cry" as a negative sentiment.**
26. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative sentiments appear. Negative words like "irritate" or "dumb" outweigh the positive words.**
27. Correct label: positive. Prediction: negative. **Reason: Review only complained about packaging but praised the product. Model outweighed the negative words like "issue" or "absurd".**
28. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative sentiments appear.**
29. Correct label: irrelevant. Prediction: positive. **Reason: Review is a very irrelevant story but the model picked up positive words like "love" and "enjoyable". Model would need to understand the context of the review/story.**
30. Correct label: negative. Prediction: positive. **Reason: The model was unable to detect sarcasm and detected more positive keywords 'world class'**
31. Correct label: positive. Prediction: negative. **Reason: Without context, the model detected more negative keywords.**
32. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**
33. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**
34. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**
35. Correct label: irrelevant. Prediction: negative. **Reason: The model only detects negative keywords. Without context it was unable to determine the text was not relevant.**
36. Correct label: negative. Prediction: positive. **Reason: Case where negative and positive sentiments exist, with negative far outweighing positive. Without context, the model detected more positive keywords.**
37. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**
38. Correct label: negative. Prediction: positive. **Reason: N-grams not used, no context to describe the positive keywords detected that refer to past sentiments.**
39. Correct label: irrelevant. Prediction: positive. **Reason: Nonsense text that is irrelevant. The model only detects positive keywords.**
40. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**

41. Correct label: negative. Prediction: positive. **Reason: Preprocessing data took away tone and sarcasm. Only positive keywords detected.**
42. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**
43. Correct label: positive. Prediction: neutral. **Reason: Negative keywords appeared in text and cancel out positive sentiment to create a neutral prediction.**
44. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**
45. Correct label: neutral. Prediction: negative. **Reason: Model only detected negative keywords such as locations.**
46. Correct label: irrelevant. Prediction: negative. **Reason: N-Grams not used. The model only detected negative keywords without context to identify that the text is not relevant.**
47. Correct label: positive. Prediction: negative. **Reason: N-Grams not used and without context, number of negative keywords outweigh positive sentiments.**
48. Correct label: neutral. Prediction: positive. **Reason: N-grams not used and only positive keywords were detected without context.**
49. Correct label: neutral. Prediction: negative. **Reason: Reason: N-grams not used and only negative keywords were detected without context.**
50. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More negative keywords appeared.**
51. Correct label: irrelevant. Prediction: negative. **Reason: N-Grams not used. Without context the model was unable to determine that the texts were not relevant and only detected negative keywords.**
52. Correct label: irrelevant. Prediction: negative. **Reason: N-Grams not used. Without context the model was unable to determine that the texts were not relevant and only detected negative keywords.**
53. Correct label: neutral. Prediction: positive. **Reason: Case where both negative and positive sentiments appear. However the model was unable to detect more positive keywords.**
54. Correct label: neutral. Prediction: positive. **Reason: Case where both negative and positive sentiments appear. However the model was unable to detect more positive keywords.**
55. Correct label: neutral. Prediction: negative. **Reason: Case where both negative and positive sentiments appear. However the model was unable to detect more negative keywords.**
56. Correct label: neutral. Prediction: negative. **Reason: Case where both negative and positive sentiments appear. However the model was unable to detect more negative keywords.**
57. Correct label: irrelevant. Prediction: positive. **Reason: N-grams was not used. Without context the model was unable to determine that the text was not relevant and only detected positive keywords.**
58. Correct label: negative. Prediction: neutral. **Reason: Case where both negative and positive sentiments appear. However the model was unable to detect more negative keywords.**
59. Correct label: neutral. Prediction: positive. **Reason: “sweet” keyword associated with positive sentiment - without context “sweet” is neutral**
60. Correct label: neutral. Prediction: positive. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More positive keywords appeared.**
61. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More negative keywords appeared.**
62. Correct label: neutral. Prediction: negative. **Reason: Case where positive and negative cancel out to create a neutral sentiment. More negative keywords appeared.**

63. Correct label: neutral. Prediction: positive. **Reason: N-gram not used, positive keyword “star” appears in text**
64. Correct label: irrelevant. Prediction: neutral. **Reason: N-gram not used, without context unable to identify that sentiments were not relevant to the product.**
65. Correct label: neutral. Prediction: positive. **Reason: N-gram not used, so positive keyword “yum” used without context**
66. Correct label: negative. Prediction: positive. **Reason: Positive keywords such as “best”, “great” detected and negative sentiments of ‘rectum, destroy inside’ not recognized**
67. Correct label: negative. Prediction: neutral. **Reason: Since n-gram was not implemented, references to locations did not provide context.**

4 – Discussion

The developed model can be used to provide a benefit in various real-world scenarios. These scenarios and the benefits will be discussed along with the realistic implications that are mostly a result of the low accuracy score of the machine learning model.

There are many uses for a natural language text sentiment analysis model on the internet. With the model that has been developed through being trained on amazon comments, sentiment can be analysed on other platforms that contain natural language text, such as twitter and Instagram. Twitter and Instagram allow for the public to comment on someone's tweet/ post. The sentiment model that has been trained on reading amazon comments can be used to review the sentiment from various twitter or Instagram post comments. For example, politicians use twitter and Instagram and other social media platforms to support their political campaign. The sentiment analysis model can be used to provide the overall sentiment of the community over specific social media posts. This technique can be used by politicians to gauge the community's sentiment over what the politician posts on social media. Whether the overall sentiment is a “3” or a “1”, the politician can draw a path forward over what the public desires. However, using this specific sentiment analysis model for something as serious as political science can cause real world implications. As previously stated, the model has an accuracy rate of 68%. Therefore, it is likely a comment can be interpreted as a sentiment of value “1” when in reality it was a neutral comment and was meant to be interpreted as a “2”. With this happening often, the overall sentiment of all the comments belonging to one social media post can be interpreted as a “1” when it should have been a “2”. This would result in the politician believing their post was bad and made a lot of people angry when the outcome was neutral. The politician would be sent in the wrong direction because they would be following the results of the inaccurate (58%) sentiment analysis machine learning model.

Another real-world implementation of our model would be labelling sentiments for nearby restaurants using Google Maps reviews. The food scene is a booming market and there are millions of reviews from self-proclaimed food-experts. The difficult part is digesting all this information for the customers and the restaurant owners. Looking at the 1–5-star overall rating of the restaurant gives a general overview but does not provide a sentiment on specific aspects of the restaurants like service or main dishes. Our model would analyse the reviews of a certain restaurant and label each of them as irrelevant, negative, neutral, and positive. This would benefit both the restaurant and the customers. The restaurant would benefit from knowing which dishes show up in the negative sentiment reviews and either make necessary changes to the dish or get rid of the dish entirely. The positive/negative sentiment would help customers when deciding which dish to try/avoid at the restaurant. This sentiment analysis will also help customers that put more emphasis on the taste and quality of the food rather than the service. If the sentiment of the negative reviews is solely based on subpar service but all of the positive sentiment is based on the positive reviews, this can help the customer when making a decision of whether or not to eat at that restaurant.

The basis of sentiment analysis relies on natural language processing on available big-data to assist in making decisions. Our model could also apply to similar review-providing platforms, such as GlassDoors, or Indeed to assist people in researching for companies that they are planning to apply for. Many of these platforms have sections that allow previous or current employees to discuss their experiences with the specific company. Sentiment analysis could be applied to these mini surveys to evaluate the company as negative, positive, or neutral, and can even be broken down to more specific topics such as work culture, compensation, safety, etc. Companies could also use this type of feedback to review their image, manage their policies and implement changes that could help with employee satisfaction, retention, recruitment and even productivity.

5 – Conclusion

After training six different machine learning models while adjusting various hyperparameters, it is conclusive that bag of words is the more appropriate feature extraction method when compared to TF-IDF. The highest accuracy score of 58.64% was achieved by using bag of words to extract the features and Naive Bayes as the classifier. Random Forest was also able to achieve the same accuracy of 58.64% when bag of words was used. This accuracy score was achieved with Naive Bayes before any tuning, which demonstrates that the hyperparameter for Naive Bayes is most optimum at the default value. The Random Forest classifier required tuning to achieve the 58.64% accuracy.

Table 5: Final accuracy score after tuning for each model examined.:

	Feature Extraction and Classifier	Accuracy Score	Hyperparameters
1	Bag of Words + Logistic Regression	0.5679012345679012	regParam: 0.3, elasticNetParam: 0.1
2	Bag of Words + Naive Bayes	0.5864197530864198	smoothing: 1.0
3	Bag of Words + Random Forest	0.5864197530864198	maxDepth: 10, maxBins: 5, numTrees: 50
4	TF-IDF + Logistic Regression	0.5679012345679012	regParam: 0.3, elasticNetParam: 0.1
5	TF-IDF + Naive Bayes	0.5617283950617284	smoothing: 1.0
6	TF-IDF + Random Forest	0.41358024691358025	maxDepth: 10, maxBins: 5, numTrees: 5

Table 5 above represents the results after tuning for each model examined. The range of accuracy of all tested combinations of feature extraction and model was between 40-60%.

Overall, a machine learning model that can review and rate amazon comments regarding sentiment has been built and implemented successfully. This model has the potential to be used for many other real-world applications such as reviewing comments on social media platforms, restaurant comments and even glass door comments. With all real-world applications, the model will pose implications due to the 58.64% accuracy score. Reviewing other feature extraction methods such as N-grams to obtain dimensional context could provide benefit to improving the model. Adjusting the preprocessing steps to include specific punctuation such as exclamation marks could also improve the model. Lastly, the various classifiers can be tuned even further with more computational cost to increase the accuracy score.

6 – References

[1] ElAsriPersonEnvelope, I., NoureddineKerzazi, GiasUddin, FoutseKhomh, Idrissi, M. A. J., Bosu, A., Murgia, A., Pang, B., Kucuktunc, O., Garcia, D., Baysal, O., Efstathiou, V., Lin, B., Imtiaz, N., Novielli, N., Calefato, F., Ahmed, T., Islam, M. R., Munezero, M. D., … Guzman, E. (2019, June 18). *An empirical study of sentiments in code reviews*. Information and Software Technology. Retrieved December 13, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S0950584919301387

[2] Gias Uddin. (n.d.). Retrieved December 13, 2022, from https://giasuddin.ca/