Personalized Restaurant Recommendation: A Cuisine Based Approach

João Victor M. Tamm[1] **, Rodrygo L. T. Santos**[1]

[1]Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)

{jvtamm, rodrygo}@dcc.ufmg.br

Abstract. *Most people use review sites such as Yelp to enhance their experiences when going out. Even though this kind of website offers a good basis, just a star review is not enough. In this paper we developed a sentiment analysis model that highlights the best and worst aspects of a few cuisines. The approach is based on a support vector machine (SVM) model that is used to unveil the sentiment tendency of each word in a review. Among the restaurant categories explored, we found that customers tend to value more service than food flavor and some common senses have been emphasized: Chinese food is usually greasy and French cuisine overpriced.*

1. Introduction

The introduction of the Web back in late 20th century is undoubtedly one of the greatest revolutions in recent human history, and therefore completely changed the way how people live but especially how they share and explain their experiences. The advent of websites which encourage users to express their opinion, such as Yelp and Trip Advisor, has greatly contributed to the growth in the amount of information available online about businesses in general. This data became very relevant and valuable for companies, business owners and for anyone who is interested in what users say about their products and how they can enhance their services.

Huge amount of data does not always guarantee a satisfactory outcome [Choi and Han, 2010], as business owners may not have time to read and absorb such amount of information. In an abundance scenario, Natural Language Processing and Sentiment Analysis have gained traction and have been the focus of major researches worldwide. In this paper we are going to use such techniques to characterize a few cuisines based on good and bad aspects reported by their consumers. What are the key factors for a restaurant success and the major customer concerns for a great meal are some of the questions that we aim to answer.

Yelp is an American company that aims to connect people with great local businesses [Yelp, 2019]. Their platform consists in a social network-like forum, where patrons can write textual reviews along with a star-rating about their experiences in all kinds of restaurants. Thousands of people use Yelp on a daily basis across the globe and recent researches have shown that the effect of these reviews on people's choice is significant. Luca [2011] ran an empirical research that showed that a one-star increase in Yelp rating leads to a 5-9 percent increase in revenue, therefore revealing how valuable Yelps data is.

The rise in popularity of websites such as Yelp draws the attention of the food industry, thus making user-provided ratings a crucial index. In previous years, important researches focused on this field have been emerging. Famous studies include an exploration of why reviews that contained the word "Groupon" provided, on average, significantly lower ratings than reviews that did not [Byers et al., 2011] and a survey that investigated the reason why users use Yelp [Hicks et al., 2012]. Being the most valuable data of Yelp's platform, exploration into the reviews using sentiment analysis and machine learning techniques must yields interesting information.

In this study, we used the numerical score given by the users to create labels for each review and then applied existing algorithms of opinion mining on the text in order to estimate how each word affects the sentiment of the customer. The analysis was made on an extensive dataset made available by the Yelp Dataset Challenge. Support Vector Machine (SVM) was the chosen machine learning algorithm and some experiments with different text cleansing and stemming and feature selection algorithms were made looking for the best classifier. A brief discussion of the algorithms tested and why some of them are not suitable for the current scenario is also made.

By the end of this research, several interesting patterns about Yelp reviews were found. Analysis on the sentiment polarity showed a preference on service over food quality, which might indicate that users do not always try foods rather than the ones they like. Aside from this, some insights that could be valuable for Yelp and new business owners were revealed.

The outline of the paper is as follows. Section 2 presents an overview of recent approaches to Sentiment Analysis, Natural Language Processing and some researches on Yelp Dataset. Section 3 describes the data and the methods used for extracting information from Yelp reviews. Section 3 contains preliminary tests and results. Finally, Section 4 wraps the research highlighting applications for the method and future studies.

2. Related Works

In this section we present the topic of Sentiment Analysis, Natural Language Processing and Machine Learning more in depth, alongside with their associated problems. Later, we finish by highlighting some ideas proposed by Yu et al. [2017].

Our present work relates to multiple lines of research. According to Medhat, Hassan, and Korashy [2014], Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward text unit. The range of applications is enormous and among them the one that is most related with this study is improving customers' relation model. SA can be classified in many levels of granularity: document classification level [Turney, 2002], sentence classification level [Hu and Liu, 2004], phrase classification level [Wilson et al., 2005] and lately word/term level [Nikos et al., 2011].

Several approaches have been used to address the sentiment hidden behind the words and expressions provided by users. Most of them are related to machine learning and many researches have been made in order to find the best one that suits NLP concerns. As the most popular are Naive Bayes (NB) [Troussas et al., 2013], Maximum Entropy (ME) [A. and Sonawane, 2016], Support Vector Machine (SVM) [Yu et al., 2017], Unsupervised Learning [Maas et al., 2011] and more recently Neural Networks [Kim, 2014]. Before the introduction of Neural Networks based methods, the state of the art was Linear SVMs an therefore this is the approach being used in this study.

Although Sentiment Analysis is an extremely powerful technique that helps business owners in the process of decision making, providing rich information given by real customers, it still faces some hurdles. Liu [2010] raises some of the main barriers concerning SA:

1. **Object identification**: This problem relates to identifying the objects to which the text is referring to. The importance of this problem arises from the fact that without knowing the object on which an opinion has been expressed, the opinion becomes useless.
2. **Feature extraction**: This problem regards extracting the main features, characteristic of the object on which the opinion has been expressed. Many attempts has been made to address this problem but this is still considered one of the main challenges encountered in SA. Currently researches are able to find nouns and noun phrases but not with much accuracy. Verb features are very common but harder to identify.
3. **Synonym grouping**: People often use different words or phrases to describe the same feature (e.g "sound" and "voice") and therefore the necessity of grouping this words or sentences exits.
4. **Opinion orientation classification**: This issue arises from the need to identify if a sentence contains or not an opinion and, if so, whether it is positive or negative. It is a problem because there are an unlimited number of expressions that people use to express opinions and, depending on the context and domain they can yield different meaning. The worst case is found when, in the same domain, the same word may indicate different opinions.

Three other major challenges facing OM are related to subjectivity and objectivity, irony and context detection. The context is so important that most SA-related problems stem from the difficulty in detecting it. Aside from that, irony is probably the most challenging barrier faced nowadays in this field as researchers do not agree completely on its nature and definition. Filatova [2012] stated that one of the major issues within the task of irony identification is the absence of an agreement among researchers (linguists, psychologists, computer scientists) on how one can formally define irony or sarcasm and their structure. The task becomes even harder as many theories that try to explain this phenomenon agree that it is impossible to come up with a formal definition of it. Likewise, it is believed that these terms are not static but undergo changes [Nunberg, 2001] and that sarcasm even has regional variations [Dress et al., 2008].

The advent of web technologies and the growth of Internet interactions alongside the increase of visual communication through emoticons, GIFs and stickers introduced new challenges for OM researchers.

Many Sentiment Analysis techniques uses support vector machines as classifier and this study is no different. Support Vector Machines is a discriminative classifier formally defined by a separating hyper plane. In other words, given labeled training data, the algorithm outputs an optimal hyper plane which categorizes new examples. In two dimensional space this hyper plane is a line dividing a plane in two parts which determine the classification of new instances.

A study that highly influenced this project was proposed by Yu et al. [2017] and consists in classifying word polarity in Yelp reviews and performing an analysis based on restaurant type. The approach is relatively simple and produces powerful results. The simplicity arises because the only data needed is a rating and a textual information per review.

3. Data and Methodology

3.1. Data description

The dataset used for this work was provided by Yelp through Yelp Dataset Challenge Round 13 and was divided into 3 JSON files which contained approximately 550.000 business basic information (name, hours, address, category, etc.), 4 million reviews and 2 million users.

For this analysis we were interested only in restaurants that have at least one category different from the ”Restaurant” one and therefore used only the business attributes and customer reviews dataset. After the filtering process we were left with 1,100,976 reviews collected from 27,793 restaurants, distributed in 13 categories across 23 American states.

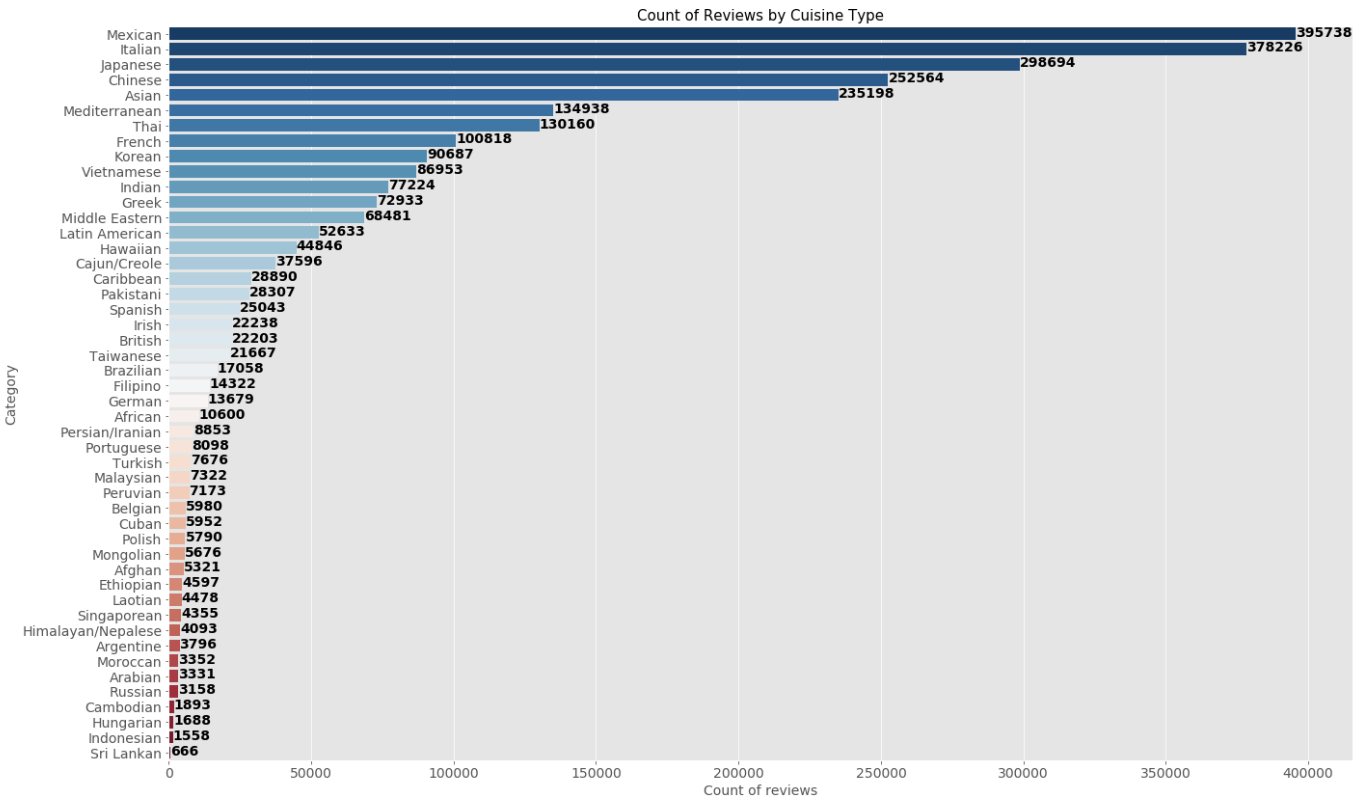

Figure 1. Distribution of reviews per categories

Attributes used for this study were just a small subset of the ones provided by the dataset. In the business data business_id was used for joining purposes and categories were used as the key for grouping restaurants reviews into their corresponding category. From the reviews side, the attributes used consisted in review_id, business_id, text content and rating. Text content was used as the corpus for the sentiment analysis and rating to create the label (positive or negative) for each review.

3.2. Data cleansing

First of all, the data that was in JSON format was converted to CSV. Both datasets were merged using the business_id attribute as a key and, for each review where the restaurant had more than one category, it was required to replicate it so that each had exactly one associated category. To reduce the amount of data, only the top 12 categories with the greatest amount of reviews were chosen. After that, reviews were labeled either positive or negative accordingly to its rating and having as threshold 3 stars. The review text was preprocessed to remove all kinds of punctuation.

Data cleansing played an important role in the experiments because the two models that were proposed differ in the way they handle the text within the review. The first one advocates that only stop words must be removed from the content. Stop words are most common words found in any natural language which carry only syntactic importance and very little or no significant semantic context in a sentence [K. and Saini, 2016]. This proposition might still leave some noise in data which could possibly lead to a lower accuracy in the method. When compared to the method described below, the dimension of the data is much higher increasing the complexity of the design.

The other approach suggested that in order to remove noise in reviews content, just positive and negative words should have been kept. Thus another dataset that consisted in a collection of positive and negative English words was needed. It was made available by Bing Liu of The University of Illinois at Chicago (UIC) at his website `https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html`. The idea behind this method is that only words that have some polarity can affect the overall opinion that is being given. It is important to highlight that a possible drawback concerns the removal of context and of words that might modify the meaning of a positive or negative term.

3.2.1. Stemming and Lemmatization

A topic in which both approaches coincide is Stemming and Lemmatization. Both methods are forms of normalization and their goal is to reduce inflectional forms of a word to a common base form, this is, to identify a canonical representative for a set of related word forms. According to Manning et al. [2008] the two techniques differ in their flavor. The authors stated that Stemming usually refers to a crude heuristic process that chops off the ends of words in the hope of achieving this goal correctly most of the time, while Lemmatization does things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base form of a word, which is known as the lemma.

Researches have shown that text normalization process highly increases the performance of NLP methods. Although both propositions are very similar and have satisfactory outcomes, Lemmatization usually yields more relevant results when compared to Stemming [Balakrishnan and Ethel, 2014]. Based on this result, the current work opted to use Lemmatization.

3.3. N-grams and Tokenization

Wikipedia defines an n-gram as a contiguous sequence of N items from a given sample of text or speech. Here an item can be a character, a word or a sentence and N can be any integer. n-grams model is one of the most widely used sentence-to-vector models since it captures the context between N-words in a sentence. In tokenization context n-grams are used to record the frequency of the word combination and how often they are likely to occur. For this study we considered only uni-grams, bi-grams and tri-grams.

Tokenization consists in the identification of each "atomic" unit of a document, this is, the task of splitting a stream of characters into words. This phase might seem simple in languages that separate words with white spaces but it could get complicated in those scenarios where this is not true such as in Chinese language. In the current research, three tokenization methods are considered: Bag of Words (BoW), Binary Bag of Words (BBoW) and Term Frequency-Inverse Document Frequency (Tf-Idf).

Bag of words is a simple method that vectorizes a text by simply counting occurrences of n-grams in a document. The result is a 1D vector in which the number of entries coincides with the size of the n-gram space. The value of each entry is exactly the amount of times the corresponding n-gram shows up in the review. Analogous do Bow, Binary Bag of Words does the same steps but instead of counting, is computes just if the n-gram appears or not, thus generating as an output a binary vector.

Term frequency-Inverse Document Frequency comes to address the problem identified with BoW that consists in giving too much emphasis in words that appear a lot in a document but do not have semantic value (e.g. the term "the" is very common so BoW will tend to incorrectly emphasize documents which happen to use the word "the" more frequently). Wikipedia defines as a numerical statistic that is intended to reflect how important a word is to a document in a collection by weighting the term frequency with the amount of documents the word shows up. The idea here is that if a word appears in many documents it should be a very common one and therefore must not have much value. The mathematical formulation used is as follows:

$$tfidf(t,d) = tf(t,d) * log\frac{N}{|\{d \in D : t \in d\}|} \tag{1}$$

where,
$tfidf(t,d)$ is the value for term t in document d,
$tf(t,d)$ is the frequency of term t in document d,
N is the total number of documents in the corpus $N = |D|$,
$\{d \in D : t \in d\}$ is the number of documents where the term t appears

3.4. Method

This study can be divided into two main parts. The first one consists in a Support Vector Machine (SVM) classifier that is used to differentiate positive and negative n-gram and to extract the respective score that let us analyze how positive or negative it really is. Further, the score results are analyzed in the point of view of the polarity of each one of the n-grams. The diagram below describes the whole process, from dataset extraction to polarity ranking.

Regarding the initial phase, two approaches were proposed and multiple combinations of tokenizers and n-gram range explored. Both models differ in the word selection

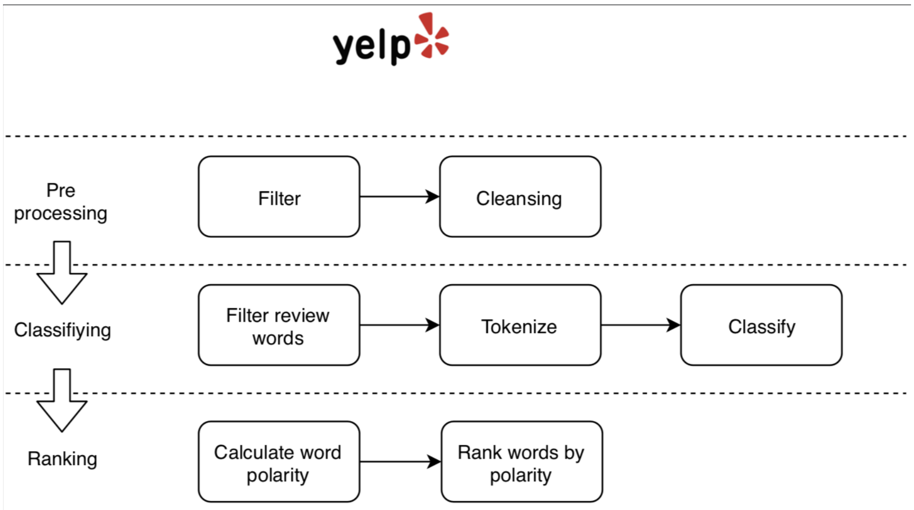

Figure 2. Model's pipeline

algorithm, while the first one opted to remove all stop words the second suggested that the best option is to only keep positive and negative words, thus eliminating any kind of noise. For each one of the models all combinations of tokenizers and n-grams cited in Subsection 3.3 were validated with 3 different categories. Data was randomly separated into training, validation and testing set according to ratio 16:4:5. To avoid over or under fitting, a cross validation was made in order to find the best hyper-parameter C. The accuracy of each classifier was evaluated using the test partition of the data. A comparison of the models was made and the model with best test performance was selected for implementing the next step.

The second phase was aimed to find customers' concerns for each of the restaurants categories. In order to do that, some vague adjectives that did not express the real sentiment of the review (e.g. wonderful, terrible, incredible, etc) were removed. Another assumption that was made is that the n-grams that were left could reflect the characteristic of different cuisines. The analysis was made base on each cuisine separately because we assumed that a word could be positive in the context of one category and negative in another (e.g. the word "sweet" could be positive in a desert category but not in the American).

To get the 'polarity score' — value that reflects the polarity of sentiment — of each cuisine, the score of each word was first multiplied by its frequency, and then normalized by the total number of reviews for the specific category. The mathematical formulation is the following:

$$polarity\_score(t, c) = score(t) * \frac{frequency(t, c)}{number\_of\_reviews(c)} \tag{2}$$

where,
$polarity\_score(t, c)$ is the factor that indicates how valuable word t is for cuisine c,
$score(t)$ is the sentiment score of word t calculated by the SVM model,
$frequency(t, c)$ is the frequency of word t among all reviews of cuisine c,

$number\_of\_reviews(c)$ is the amount of reviews in cuisine (c).

The polarity score we calculated, basically shows how much a word contributes to the score of all restaurants in a certain cuisine.

4. Results and Discussions

High accuracy was obtained by both methods proposed, reaching marks above 90%. Analyzing the cross validation results, we discovered that, on average, the results with best accuracy were achieved when the hyper-parameter C was set to 0.1.

Setting C to 0.1 and running SVM classifier on the test data, the accuracy obtained by the stop word approach was nearly 95% while the positive/negative words method got a little bit over 92%. We also find that Bag of Words (BoW) using uni-grams was the design that had the best cost benefit. Although Term Frequency-Inverse Document Frequency usually outperforms BoW in most of the cases, we think that in our study the results turned out to be different because we removed the stop words and in one of the cases we kept just relevant words and therefore not having the issue of words appearing very often in most documents.

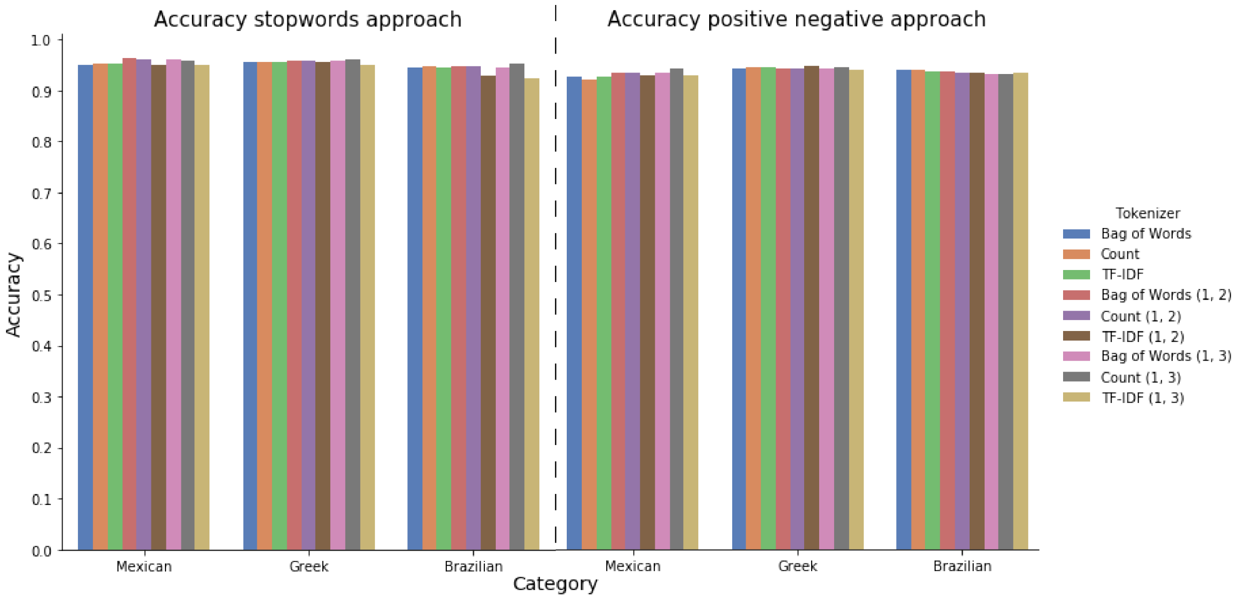

Figure 3. Model's accuracy comparison

Although the stop words approach outperformed the positive negative one, its complexity was extremely higher and we lacked the computational resources required to move with it. The dimensionality of the features turned out to be extremely high, making the process of calculating terms frequency nearly impossible with the computer we had available. Therefore, we chose to go with the positive negative approach, leaving the stop words one for a future research.

Moving on from the classification side, results regarding polarity ranking were quite unexpected. It seems that customers value service over food flavors, since words such as *friendly* are top ranked. Another factor that might indicates that service is a crucial factor is that words like *fast* is always ranked in the top 10 positive words. One possible explanation of this coincidence is that when customers decide where to have a meal, they would usually choose the specific kinds of cuisine they prefer. As for the food flavor customers highlighted the *freshness* and *tenderness* as the most valuable aspects.

Category	Top 5 Positive Words				
Mexican	friendly	fresh	fast	authentic	tender
Italian	friendly	fresh	reasonable	attentive	authentic
Japanese	friendly	fresh	attentive	reasonable	fast
Chinese	friendly	fresh	authentic	fast	reasonable
Asian	friendly	fresh	reasonable	variety	clean
Mediterranean	friendly	fresh	healthy	tender	reasonable
Thai	friendly	fresh	reasonable	hot	clean
French	friendly	fresh	sweet	impeccable	tender
Korean	friendly	tender	fresh	reasonable	attentive
Vietnamese	friendly	fresh	reasonable	clean	fast
Indian	friendly	fresh	reasonable	variety	authentic
Greek	friendly	fresh	generous	tender	healthy
Brazilian	friendly	fresh	tender	attentive	fun

Figure 4. Top positive polarity score words by cuisine

Category	Top 5 Negative Words				
Mexican	bland	cold	slow	hard	expensive
Italian	cold	hard	bland	slow	expensive
Japanese	slow	hard	cold	bland	expensive
Chinese	bland	cold	hard	sour	greasy
Asian	bland	cold	slow	expensive	hard
Mediterranean	cold	bland	hard	slow	expensive
Thai	bland	sour	slow	cold	greasy
French	cold	slow	hard	bland	expensive
Korean	bland	slow	cold	expensive	mediocre
Vietnamese	bland	slow	expensive	greasy	mediocre
Indian	bland	cold	expensive	sweet	hard
Greek	cold	bland	slow	expensive	sad
Brazilian	cold	slow	seasoned	hard	tough

Figure 5. Top negative polarity score words by cuisine

From the negative words list, we could observe that the blandness, coldness and hardness of the food are more likely to to be the reason for a low score. A good recommendation for people who runs business within these cuisines is to focus on a better seasoning, food temperature and to not overcook the food, as customers complain a lot about this. An interesting fact is that people who go to Brazilian restaurants complained about food seasoning, which might indicates a cultural aspect, since this cuisine is much distinct from the other ones. The appearance of the word *slow* in different categories support our hypothesis that service is the main key factor for a good experience when going out.

A really interesting information that we were able to identify is related to the customers' perception about pricing. In the positive words rank, many users referred to the word reasonable to characterize a cuisine, which might indicates that the price is fair relative to the experience they had in the restaurant. As for the negative words list, many customers highlighted that the food was expensive and overpriced.

Finally, probably the most expected result but still very interesting is that we were able to obtain results that match exactly with the common sense. Japanese cuisine offers many raw dishes, so it was expected and was confirmed that freshness is definitely a key factor for the restaurant success. Our study also showed that customers' opinions have emphasized the reputation of French bistros for being expensive and over-priced. As for the last common sense, we would like to highlight the greasiness of the Chinese cuisine, that is very famous for its deep fried foods and sweet sauces.

5. Conclusion an Future Works

In this paper, we implemented an efficient SVM model to calculate the relevance of the sentiment expressed by words in reviews. Other than extracting keywords, this model could be used, with few modifications, to predict ratings of Yelp's tips, thus giving more reasonable overall ratings for restaurants, since tips are highly valuable for business owners. Changing the context from cuisines to actual restaurants is also something that could yield interesting results and probably an even powerful decision making helper for stakeholders.

Taking our analysis into consideration is pretty easy to create rich recommendations for each cuisine. The features that this method helps to extract from any set of reviews can be used by restaurant owners for essential decision making. Without having to read all the reviews they can understand why customers like or dislike their establishment and even compare his services with the ones offered by their competitors. A possible drawback that must be highlighted concerns the amount of reviews required in order to have a satisfactory outcome.

From the standpoint of Yelp and its customers, the method can be used as a new feature in the Yelp Dashboard to help users choose the location that best suits their needs. The panel could show the key positive and negative feelings drawn from the ratings to provide users with a description of the main characteristics of the venue, thus helping them in the choice process

Other application of the method may include a consulting process that helps people who wants to start a local business in the food industry. Suppose someone would like to open a Japanese and Asian restaurant, by operating the steps described in this research he would definitely be more accurate in his decision process if it is worth or not. By knowing what are the points in which similar restaurants are failing to deliver to their customers and the ones they are delivering with excellence, the new business owner can just replicate the good stuff and enhance the bad ones.

Although the performance of the model used is satisfactory, there are plenty of space for improvement. One of the suggestions for future works is to try to execute the whole process with the stop words approach and maybe explore if this method also applies for Yelp's Tips. In addition, using the extracted words as features for a recommendation system is also something we would like to explore in the near future. In order to have a more robust system this recommendation could easily take into consideration compliments (e.g. funny, cool, useful, etc.) given the reviews by other patrons.

References

Vishal A. and S.S. Sonawane. Sentiment analysis of twitter data: A survey of techniques. *International Journal of Computer Applications*, 139(11):5–15, April 2016.

ISSN 0975-8887. doi: 10.5120/ijca2016908625. URL `http://dx.doi.org/10.5120/ijca2016908625`.

Vimala Balakrishnan and Lloyd-Yemoh Ethel. Stemming and lemmatization: A comparison of retrieval performances. *Lecture Notes on Software Engineering*, 2:262–267, 1 2014. doi: 10.7763/LNSE.2014.V2.134.

John W. Byers, Michael Mitzenmacher, and Georgios Zervas. Daily deals: Prediction, social diffusion, and reputational ramifications, 2011.

Sang-Min Choi and Yo-Sub Han. A content recommendation system based on category correlations. In *Fifth International Multi-conference on Computing in the Global Information Technology*, 2010.

Megan L. Dress, Roger J. Kreuz, Kristen E. Link, and Gina M. Caucci. Regional variation in the use of sarcasm. *Journal of Language and Social Psychology*, 27(1):71–85, 2008. doi: 10.1177/0261927X07309512. URL `https://doi.org/10.1177/0261927X07309512`.

Elena Filatova. Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 392–398, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/661_Paper.pdf`.

Amy Hicks, Stephen Comp, Jeannie Horovitz, Madeline Hovarter, Maya Miki, and Jennifer Bevan. Why people use yelp.com: An exploration of uses and gratifications. *Computers in Human Behavior*, 28:2274–2279, 11 2012. doi: 10.1016/j.chb.2012.06.034.

Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA, 2004. ACM. ISBN 1-58113-888-1. doi: 10.1145/1014052.1014073. URL `http://doi.acm.org/10.1145/1014052.1014073`.

Jaideepsinh K. and Jatinderkumar Saini. Stop-word removal algorithm and its implementation for sanskrit language. *International Journal of Computer Applications*, 150: 15–17, 9 2016. doi: 10.5120/ijca2016911462.

Yoon Kim. Convolutional neural networks for sentence classification, 2014.

Bing Liu. Sentiment analysis: A multi-faceted problem. *IEEE Intell Syst*, 25, 1 2010.

Michael Luca. Reviews, reputation, and revenue: The case of yelp.com. *SSRN Electronic Journal*, 9 2011. doi: 10.2139/ssrn.1928601.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 142–150, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. ISBN 978-1-932432-87-9. URL `http://dl.acm.org/citation.cfm?id=2002472.2002491`.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008. ISBN 0521865719, 9780521865715.

Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4):1093–1113, 2014. ISSN 2090-4479. doi: https://doi.org/10.1016/j.asej.2014.04.011. URL `http://www.sciencedirect.com/science/article/pii/S2090447914000550`.

G. Nunberg. *The Way We Talk Now: Commentaries on Language and Culture from NPR's "Fresh Air"*. Houghton Mifflin, 2001. ISBN 9780618116027. URL https://books.google.com.br/books?id=oEliAAAAMAAJ.

C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro. Sentiment analysis of facebook statuses using naive bayes classifier for language learning. In *IISA 2013*, pages 1–6, July 2013. doi: 10.1109/IISA.2013.6623713.

Peter D. Turney. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073153. URL https://doi.org/10.3115/1073083.1073153.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 347–354, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/H05-1044.

Yelp. Yelp. https://www.yelp.com, 2019. Accessed: 2019-08-15.

Boya Yu, Jiaxu Zhou, Yi Zhang, and Yunong Cao. Identifying restaurant features via sentiment analysis on yelp reviews, 2017.