

Review Analysis

(Engineering Project)

Sachin Devangan, Vrund Patel

sdevanga@stevens.edu vpatel22@stevens.edu

Department of Computer Science

Stevens Institute of Technology

Abstract

Foretelling the value of product appraisals is a crucial element of various ecommerce endeavors, including review positioning and suggestion. Prior research such as The Bag of Words (BoW) model which is one of the primary examples where the model was widely praised for its simplicity and efficiency. However, previous works tend to destroy the original natural language, disrupts syntactic structures, and ignores word order as they were all evaluated on pre-defined datasets. As a result, various alternative methods such as higher order n-grams were introduced. In this project, we have initiated a unique effort by reviewing analysis on real-time datasets where we will get insights of the reviews from every available website. The objective is to fetch the available reviews for any business, and then process it through our model to analyze its polarity or sentiment as well as subjectivity. Moreover, the model also extracts keywords from the reviews. In the context of review analysis, keyword extraction has been used for identifying common themes or topics in customer reviews.

Introduction

The Internet has had a profound impact on our lives, transforming the way we communicate, access information, and conduct business. One of the most significant ways in which the Internet has affected our lives is through the ability to access vast amounts of information at our fingertips. With a few clicks, we can find answers to almost any question, learn about any topic, and access a wealth of knowledge that was once only available to a select few. This has revolutionized education and made it possible for anyone, regardless of their location or circumstances, to learn and grow.

The Internet can be a valuable resource for product reviews, as it allows individuals to easily access a wide range of opinions and experiences from other customers. By reading reviews of a product before making a purchase, consumers can gain insight into the quality and functionality of the product, as well as any potential issues or problems they may encounter. This can help them to make informed decisions and choose the best product for their needs.

Furthermore, the Internet allows for the creation of online review platforms, where consumers can easily leave reviews and ratings for products they have purchased. This can help to build trust and credibility with potential customers, as they can see the

experiences of others before making a purchase decision. Overall, the Internet is a powerful tool for product reviews, allowing consumers to make informed decisions and helping businesses to improve and grow.

Product reviews have had a profound impact on the way we make purchasing decisions, as they provide valuable insights into the quality and functionality of products. With the ability to easily access a wide range of opinions and experiences from other customers online, consumers can make informed decisions and choose the best product for their needs.

In addition to helping consumers, product reviews can also benefit companies and organizations. By analyzing the sentiment of product reviews, businesses can identify common issues or problems faced by customers and use this information to make improvements to their products or services. This can lead to increased customer satisfaction and loyalty and a better reputation for the company.

Online review platforms also allow consumers to easily leave reviews and ratings for products they have purchased, helping to build trust and credibility with potential customers. By seeing the experiences of others before making a purchase decision, customers can make informed decisions and feel confident in their purchases. Overall, product reviews are a powerful tool for both consumers and businesses, allowing for informed decision making and continuous improvement.

Related Work

There has been a significant amount of research in the field of Natural Language

Processing (NLP) on Sentiment Analysis, which covers everything from classifying the sentiment of an entire document [1] to classifying the sentiment of individual words and phrases [2]. Hatzivassiloglou and McKeown have made contributions to the field by developing methods for extracting semantic information from large corpora.

The Bag of Words (BoW) model was initially implemented to map feature vectors to documents and subsequently segment them using machine learning techniques. This approach was widely praised for its simplicity and efficiency. However, a major disadvantage of this method is that it destroys the original natural language, disrupts syntactic structures, and ignores word order. As a result, various alternative methods such as higher order n-grams were introduced [3-6] in order to address this issue.

As mentioned in [7], part of speech tagging is a technique used in syntactic analysis that is referred to as surface forms because it relies on text patterns [10-14] for tactical [8] and lexical information [9] rather than semantic meaning. Sentiment analysis also involves adding word polarity to the previously described features, using sentiment lexicons [15] to gather information. This method has several advantages [16], such as the ability to take into account negation and intensifiers, as well as the valence shift of sentiment in linguistic content. Additionally, it can differentiate the sentiment orientation of lexical entities based on their characteristics. However, lexicon-based approaches have several drawbacks [17-20], including the need for a reliable and consistent lexicon and the availability of options for opinion words that vary by language, context, and domain. This makes it challenging to maintain independent lexicons. In a machine learning-driven approach, fundamental questions such

as selecting a classification algorithm, determining relevant features, and extracting features from text must be addressed. In the past, manual feature engineering was used, which was time-consuming. Deep learning techniques can be used as an alternative to this approach.

Our Work

There are several models that utilize pre-defined datasets to perform sentiment analysis on reviews. However, this approach can be limiting as it restricts the model to the data it was trained on. In contrast, our proposed method involves collecting reviews from active websites and creating a dataset from them, enabling us to effectively train the model and predict the sentiment of reviews based on the specific product being evaluated. Additionally, our model incorporates features such as keyword extraction and machine translation to improve insights and user experience.

We will utilize datasets (composed of review sections) from websites that adopt a business model centered on providing customer satisfaction through product reviews. To gather these reviews, we will utilize the requests command and pull all reviews related to a specific product, creating a dataset from the collected data to analyze it with our model and extract insights. This approach allows us to not be limited to any dataset, enabling us to apply our model to a wide range of review-based businesses and products. The collected reviews will be converted into text format and then analyzed using our review model.

Methodology

1. Getting Dataset

As not bound by predefined dataset, we created a small dataset with the help of data scraping from an online website called Yelp.

What is Data Scraping?

Data scraping, also known as web scraping or web data extraction, is the process of extracting data from websites or other online sources using automated tools. Data scraping can be used to gather large amounts of data quickly and efficiently, without the need to manually visit each website or enter data manually.

There are various tools and techniques that can be used for data scraping, including web scraping libraries and frameworks in programming languages like Python, Java, and Ruby, as well as standalone web scraping tools and browser extensions.

To perform data scraping, you typically need to follow these steps:

1. Identify the website or online source from which you want to extract data.
2. Inspect the HTML or other source code of the website to identify the structure and layout of the data you want to extract.
3. Write a script or use a tool to send HTTP requests to the website and retrieve the HTML or other source code.
4. Parse the source code using a library or tool to extract the data you are interested in.
5. Clean and process the data as needed, and store it in a format that is convenient for analysis or further processing.

There are various considerations to keep in mind when performing data scraping, including:

1. Web scraping can be considered a form of web crawling, and it is

subject to the terms of service and policies of the websites you are scraping. It is important to respect the terms of service and policies of the websites you are scraping, and to avoid scraping sensitive or protected data.

2. Data scraping can be time-consuming and resource-intensive, especially if you are extracting large amounts of data from multiple websites. It is important to optimize your scraping process to minimize the time and resources required.
3. The structure and layout of websites can change over time, which can affect the accuracy and reliability of your data scraping process. It is important to periodically check and update your data scraping scripts or tools to ensure that they are still working correctly.

Overall, data scraping can be a powerful tool for gathering and analyzing data from online sources. It can provide valuable insights and help organizations make informed decisions based on real-time data. However, it is important to use data scraping responsibly and ethically, and to respect the terms of service and policies of the websites you are scraping.

How did we scrape data?

Yelp is a website that publishes crowd-sourced reviews about business. We used the 'requests' library in python to request data from the website. The requests library is a powerful tool for interacting with web servers and APIs, and it is widely used in Python for web scraping, data analysis, and other tasks that involve retrieving data from the internet. Since the website has many business, we selected a random business i.e. Salmon Bar (a Japanese restaurant in California).

Later we used Beautiful soup library to parse html data from the site and only extract the class of the web page that was useful for us, i.e the class containing all customer reviews. BeautifulSoup is a Python library for parsing and navigating HTML and XML documents. BeautifulSoup is a useful tool for web scraping and data analysis tasks that involve parsing HTML or XML documents. It provides a simple and intuitive interface for navigating and extracting data from these documents.

The findall() method is a method of the re (regular expression) module in Python, which is part of the Python standard library. It is used to search a string for all occurrences of a pattern specified by a regular expression.

To use the findall() method, you first need to import the re module and compile a regular expression pattern using the re.compile() function. Then you can call the findall() method on a string, passing the compiled regular expression as an argument.

2. Data Cleaning and Insights

Why do we need data cleaning?

Data cleaning is the process of identifying and correcting or removing incorrect, incomplete, or irrelevant data from a dataset. Data cleaning is a crucial step in the data preparation process, as it ensures that the data is accurate, consistent, and ready for analysis or other purposes.

There are various techniques and approaches that can be used for data cleaning, depending on the nature and quality of the data. Some common tasks involved in data cleaning include:

1. Detecting and correcting errors or inconsistencies in the data, such as spelling mistakes, typos, or incorrect values.
2. Identifying and removing duplicate data or records.
3. Removing or imputing missing values or null values.
4. Handling or converting data in different formats or types, such as dates or numbers.
5. Standardizing or normalizing the data, such as converting all values to a common format or scale.
6. Filtering or selecting relevant data for the specific purpose or analysis.

To perform data cleaning, you typically need to follow these steps:

1. Inspect and explore the data to understand its structure, quality, and any issues or problems.
2. Define the cleaning tasks and procedures based on the specific issues or problems identified in the data.
3. Write or use scripts or tools to perform the data cleaning tasks.
4. Verify and validate the cleaned data to ensure that it is accurate and complete.
5. Store the cleaned data in a format that is convenient for further analysis or processing.

There are various considerations to keep in mind when performing data cleaning, including:

1. Data cleaning can be a time-consuming and resource-intensive process, especially if the data is large or complex. It is important to optimize your data cleaning process to minimize the time and resources required.

2. The quality and reliability of the cleaned data depend on the accuracy and completeness of the cleaning tasks and procedures. It is important to carefully plan and document the data cleaning process to ensure that it is effective and thorough.
3. The data cleaning process can involve making subjective or judgmental decisions about how to handle certain issues or problems in the data. It is important to be consistent and transparent in these decisions and to document them clearly.

Overall, data cleaning is a crucial step in the data preparation process that ensures that the data is accurate, consistent, and ready for analysis or other purposes. Effective data cleaning requires careful planning and attention to detail, as well as the use of appropriate tools and techniques.

During our step of data cleaning and getting insights, we got the count of all the words, characters, average words in a particular review. We also counted the stopwords in each review so that the cleaning process becomes easier. Later, we converted each review to lower case letters, removed all the punctuations, stopwords, and all the common words so that we are left with only the raw data that is required for analysis.

For these, we used the help of nltk library. The Natural Language Toolkit (nltk) is a Python library for working with human language data (text). It provides a wide range of tools and resources for natural language processing (NLP), including tools for tokenizing, stemming, lemmatizing, tagging, parsing, and more.

3. Sentiment Analysis, Keyword Extraction, Machine Translation.

For this part of the model, the TextBlob library was used to get sentiment polarity and

subjectivity from each review. The polarity metric refers to the degree to which the text analyzed is positive or negative, between a range of -1 to 1. A score of 1 means highly positive whereas -1 is considered well and truly negative. We also analyzed subjectivity, this is the degree to which the text analyzed relates to personal emotion or factual information between a scale of 0 to 1. With scores closer to one indicating a higher level of subjectivity and being based mostly on opinion. Feeding the cleaned data to the library we got, polarity and subjectivity of reviews ranging from -1 to 1 and 0 to 1 respectively.

From the reviews of our model, we could see that the reviews were mostly positive, which matched the data on the actual website.

For Keyword extraction, we use rake library that helps in the extraction. The RAKE algorithm works by dividing the text into individual tokens (words, punctuation marks, etc.) and then identifying the most important words or phrases based on their frequency and their degree of uniqueness within the text.

To use the RAKE algorithm in the nltk library, you need to first install the nltk library and then import the RAKE class from the `nltk.rake.rake` module. Then you can create a RAKE object, passing a list of stop words (common words that should be ignored) as an argument.

Next, you can call the `extract_keywords_from_text()` method on the RAKE object, passing a piece of text as an argument. This will return a list of tuples, each containing a keyword and its associated score.

The RAKE algorithm can be useful for tasks such as keyword extraction, document summarization, and information retrieval, where it is important to identify the most

important or relevant words or phrases in a piece of text. However, it is important to keep in mind that the results of the RAKE algorithm may vary depending on the specific text and the stop words used.

For the machine translation part, a translation pipeline was used. A translation pipeline is a sequence of steps or processes that are used to translate a piece of text from one language to another.

The specific steps or processes involved in a translation pipeline can vary depending on the tools and techniques used, but a typical translation pipeline might include the following steps:

1. Tokenization: This step involves breaking the input text into individual tokens (words, punctuation marks, etc.), so that they can be processed and translated individually.
2. Part-of-speech tagging: This step involves identifying the part of speech (noun, verb, adjective, etc.) of each token in the input text, which can be used to disambiguate the meaning of the token and choose the appropriate translation.
3. Lexical translation: This step involves translating individual tokens or phrases from the input language to the target language using a lexical resource, such as a translation dictionary or machine translation model.
4. Syntactic translation: This step involves reordering and modifying the structure of the translated text to match the syntax of the target language. This can involve techniques such as inserting or deleting words, changing the order of words, or adding inflections or other grammatical elements.
5. Post-processing: This step involves cleaning and formatting the translated

text to make it more readable and appropriate for the intended audience. This can involve tasks such as capitalizing the first letter of a sentence, correcting spelling mistakes, or adding punctuation marks.

Overall, a translation pipeline is a set of steps or processes that are used to translate a piece of text from one language to another, using a combination of lexical and syntactic techniques. The specific steps or processes involved in a translation pipeline can vary depending on the tools and techniques used, but they generally involve breaking the text into tokens, translating the tokens or phrases, and reordering and modifying the structure of the translated text to match the syntax of the target language.

Future Work

In our research, we utilize simple techniques to translate specific text, in this case keywords, into different languages using the transformer library. This approach has the potential to improve the efficiency and effectiveness of machine translation for a variety of purposes.

There are several areas that could be explored as part of future work. Some possible directions for future work could include:

1. Improving the accuracy and reliability of the review analysis: This could involve developing or fine-tuning machine learning models or other algorithms to better identify and classify the sentiment of reviews, or developing techniques to handle specific types of text or languages.
2. Expanding the scope and coverage of the review analysis: This could involve collecting and analyzing reviews from a wider range of sources or languages, or developing

techniques to handle different types of reviews (such as product reviews, movie reviews, etc.).

3. Enhancing the visualization and presentation of the review analysis results: This could involve developing interactive or dynamic visualizations to better communicate the insights and trends in the review data, or integrating the review analysis with other tools or platforms to make the results more accessible and useful to a wider audience.
4. Incorporating additional data sources or features: This could involve integrating the review analysis with other data sources or features, such as demographic data, customer behavior data, or social media data, to provide a more comprehensive and nuanced understanding of the reviews and the customers who wrote them.

Overall, there are many potential directions for future work in a review analysis project, depending on the specific goals and needs of the project. Further research and development in these areas could help to improve the accuracy, coverage, and usefulness of the review analysis and provide valuable insights for businesses and organizations.

References

- [1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/15000000011.
- [2] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the eighth conference on European chapter of the*

Association for Computational Linguistics -, 1997, pp. 174–181, doi: 10.3115/979617.979640.

[3] Vijayakumar, T., Mr R. Vinothkanna, and M. Duraipandian. "Fusion based Feature Extraction Analysis of ECG Signal Interpretation—A Systematic Approach." *Journal of Artificial Intelligence* 3, no. 01 (2021): 1-16.

[4] Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data*, 5(1), 1-25.

[5] Chakrabarty, Navoneel, and Sanket Biswas. "Navo Minority Over-sampling Technique (NMOTe): A Consistent Performance Booster on Imbalanced Datasets." *Journal of Electronics* 2, no. 02 (2020): 96-136.

[6] Chakraborty, K., Bhattacharyya, S., Bag, R., & Hassanien, A. E. (2018, February). Comparative sentiment analysis on a set of movie reviews using a deep learning approach. In the *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 311-318). Springer, Cham.

[7] Hariharakrishnan, Jayaram, and N. Bhalaji. "Adaptability Analysis of 6LoWPAN and RPL for Healthcare applications of Internet-of-Things." *Journal of ISMAC* 3, no. 02 (2021): 69-81.

[8] Hanafy, M., Khalil, M. I., & Abbas, H. M. (2018, September). Combining classical and deep learning methods for twitter sentiment

analysis. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition* (pp. 281-292). Springer, Cham.

[9] Chen, Joy long Zong, and P. Hengjinda. "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method-A Comparative Study." *Journal of Artificial Intelligence* 3, no. 01 (2021): 17-33.

[10] Pasupa, K., & Ayutthaya, T. S. N. (2021). Hybrid deep learning models for Thai sentiment analysis. *Cognitive Computation*, 1-27.

[11] Haoxiang, Wang, and S. Smys. "Overview of Configuring Adaptive Activation Functions for Deep Neural Networks-A Comparative Study." *Journal of Ubiquitous Computing and Communication Technologies (UCCT)* 3, no. 01 (2021): 10-22.

[12] Ranganathan, G. "A Study to Find Facts Behind Preprocessing on Deep Learning Algorithms." *Journal of Innovative Image Processing (JIIP)* 3, no. 01 (2021): 66-74.

[13] Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications*, 118, 272-299.

[14] Smys, S., and Jennifer S. Raj. "Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study." *Journal of trends in Computer Science and Smart technology (TCSST)* 3, no. 01 (2021): 24-39.

[15] Adam, Edriss Eisa Babikir. "Survey on Medical Imaging of Electrical Impedance Tomography (EIT) by Variable Current Pattern Methods." Journal of ISMAC 3, no. 02 (2021): 82-95.

[16] Joe, Mr C. Vijesh, and Jennifer S. Raj. "Location-based Orientation Context Dependent Recommender System for Users." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 14-23.

[17] Shirley D.R.A., Sundari V.K., Sheeba T.B., Rani S.S. (2021) Analysis of IoT Enabled Intelligent Detection and Prevention System for Drunken and Juvenile Drive Classification. In: Kathires M., Neelaveni R. (eds) Automotive Embedded Systems. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-59897-6_10

[18] Chourasia, Mayank, Shriya Haral, Srushti Bhatkar, and Smita Kulkarni. "Emotion recognition from speech signals using deep learning." Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020 (2021): 471-481.

[19] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment Analysis

of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. Applied Soft Computing, 97, 106754.

[20] Janoria, Honey, Jasmine Minj, and Pooja Patre. "Classification of Skin Disease Using Traditional Machine Learning and Deep Learning Approach: A Review." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 643-651. Springer Singapore, 2021.

[21] Request- <https://requests.readthedocs.io/en/latest/>

[22] Beautiful Soup - <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

[23] Nltk - <https://www.nltk.org/>

[24] Textblob - <https://textblob.readthedocs.io/en/dev/>

[25] Rake - <https://csurfer.github.io/rake-nltk/build/html/index.html>

[26] Transformer - <https://huggingface.co/docs/transformers/index>

Appendix

```
[ ] for review in reviews:
    print(review, '\n')
```

Very good food and excellent service!! We were driving around the area and saw a grand opening sign so we decided to give it a try. So glad we did! We ordered the lunch special Attended the soft opening and experience was a vibe. Great music, soft atmosphere, the smoked salmon was worth the wait and they have a fruit tea that will set your taste buds The salmon is really wonderful! Packed with flavor, juicy, perfectly cooked. I tried the bacon salmon and it was such a unique combination - smokey and teriyaki like. I thought Fabulous, friendly and sometimes witty service, delicious food stylishly presented and a reasonable price with an amiable ambiance...so that's a yes, we luv'd this place and ku 4.5 STARSWhat to do when you go to Salmon Bar?Obvs, get salmon wasted!Salmon Bar is a new salmon-centric restaurant. They've taken over Krua Thai which has moved in with Raw Su Taste: 4.5/5Service: 5/5Value: 4/5Happy Hour Value: 5/5Ordered the Salmon Volcano, Lion King, Salmon Nachos, Clam Chowder Croquettes, Oysters. I like the focus on salmon and ho Can you build a restaurant around salmon? Yes, you can!If the parking gods deem you lucky this is a great place to escape the hustle and bustle of the mission.With the rising p We were wandering around SF for the day before seeing a dragapella show nears Hays Valley and when we drove past this oddly named restaurant with a banner that said "Soft Openi So my partner and I decided to stop in for the soft open and happened to wander in during happy hour which was a pleasant surprise. This place is a salmon heaven without a doub Nice place , nice vibe good wine and food What else can I say hahaI don't have enough words it's just good

Fig.1 extracted reviews from the website

```
[ ] df.head()
```

	review	word_count	char_count	avg_word	stopword_coun	review_lower	review_nopunc	review_nopunc_nostop	review_nopunc_nostop_nocommon
0	Very good food and excellent service!! We were...	47	241	4.148936	17	very good food and excellent service!! we were...	very good food and excellent service we were d...	good food excellent service driving around are...	good food excellent service driving around are...
1	Attended the soft opening and experience was a...	39	204	4.256410	16	attended the soft opening and experience was a...	attended the soft opening and experience was a...	attended soft opening experience vibe great mu...	attended soft opening experience vibe great mu...
2	The salmon is really wonderful! Packed with fl...	79	442	4.607595	31	the salmon is really wonderful! packed with fl...	the salmon is really wonderful packed with fla...	salmon really wonderful packed flavor juicy pe...	salmon really wonderful packed flavor juicy pe...
3	Fabulous, friendly and sometimes witty service...	42	264	5.309524	13	fabulous, friendly and sometimes witty service...	fabulous friendly and sometimes witty service ...	fabulous friendly sometimes witty service deli...	fabulous friendly sometimes witty service deli...
4	4.5 STARSWhat to do when you go to Salmon Bar?...	729	4198	4.759945	217	4.5 starswhat to do when you go to salmon bar?...	45 starswhat to do when you go to salmon barob...	45 starswhat go salmon barobvs get salmon wast...	45 starswhat salmon barobvs salmon wastedsalmo...

Fig.2 data after performing cleaning and drawing insights

	review	polarity		review	subjectivity
0	Very good food and excellent service!! We were...	0.509524	0	Very good food and excellent service!! We were...	0.695238
1	Attended the soft opening and experience was a...	0.466667	1	Attended the soft opening and experience was a...	0.475000
2	The salmon is really wonderful! Packed with fl...	0.555208	2	The salmon is really wonderful! Packed with fl...	0.779167
3	Fabulous, friendly and sometimes witty service...	0.434375	3	Fabulous, friendly and sometimes witty service...	0.744643
4	4.5 STARSWhat to do when you go to Salmon Bar?...	0.152676	4	4.5 STARSWhat to do when you go to Salmon Bar?...	0.481609
5	Taste: 4.5/5Service: 5/5Value: 4/5Happy Hour V...	0.187500	5	Taste: 4.5/5Service: 5/5Value: 4/5Happy Hour V...	0.500000
6	Can you build a restaurant around salmon? Yes,...	0.519048	6	Can you build a restaurant around salmon? Yes,...	0.705060
7	We were wandering around SF for the day before...	0.137314	7	We were wandering around SF for the day before...	0.535119
8	So my partner and I decided to stop in for the...	0.348431	8	So my partner and I decided to stop in for the...	0.625920
9	Nice place , nice vibe good wine and food What...	0.520000	9	Nice place , nice vibe good wine and food What...	0.740000

Fig.3 polarity and subjectivity values

```
[['grand opening sign', 'excellent service !!', 'miso soup'],
['good way lol ).', 'taste buds', 'soft opening'],
['unusual pairing', 'unique combination', 'teriyaki like'],
['delicious food stylishly presented',
'sometimes witty service',
'better sf hh'],
['4 ), chicken karaage ($ 4 ), gyoza ($ 4 ),',
'dozen oysters ($ 18 ), housemade ankimo ($ 8 ), yakitori',
'glass ($ 5 -$ 7 ). f'],
['would come back', 'create unique dishes', 'clam chowder croquettes'],
['parking gods deem', 'shish salmon leaf', 'restaurant around salmon'],
['nicely fried korean style boneless chicken thighs',
'seafood forward japanese style restaurant included',
'divine ($ 8 ), shrimp wrapped'],
['crispy crunchy outer layer',
'second time trying oysters',
'really tasty complimentary items'],
['nice vibe good wine', 'nice place', 'say hahai']]
```

Fig. 4 Keywords extracted from each review

✓ 0s	[50] results	[{'translation_text': 'zurückkehren'}, {'translation_text': 'einzigartige Gerichte zu kreieren'}, {'translation_text': 'clam chowder croquettes'}]
✓ 0s	[51] keywords[5]	['would come back', 'create unique dishes', 'clam chowder croquettes']

Fig.5 Machine Translation performed on 5th review