Aspect extraction from online product reviews

Guido Baels

University of Twente

ABSTRACT

Nowadays, online reviews are of great value to a customer, since they can indicate how valuable a product can be. To help a customer gain insight on the information in the reviews it would be beneficial to train a model that can extract aspects of the reviews automatically. In this way, it can quickly be seen if a product is of good quality. The same goes for companies, who quickly want to see how their products are being valued. That is why research needs to be done on the problems that arise when creating such automated models. The contribution of this paper is that it has shown that even though creating a supervised model is more time-consuming than an unsupervised model, the results in the end are worth the time that it takes to annotate data. This is because the unsupervised model has shown to be way worse at mining aspects from laptop reviews.

1. INTRODUCTION

Over the past years, online reviews have increased in numbers and importance [9, 10]. They have become more important since they form the most reliable way for customers to gain an insight into the value of an online product before they decide to buy it [18]. This means that favourable reviews can incentivise customers to purchase a product and vice versa [6]. Next to that, it is also useful for companies to monitor reviews to see how they can improve their product so they can increase customer satisfaction[17]. Of course, the companies can already easily see the star rating of reviews, but this does not always give enough information on how to improve the product to increase customer satisfaction. If the companies use aspect analysis with sentiment analysis they can see specifically what part of the product customers are complaining about and use this to improve their product or service. This shows how research on aspect extraction is of great importance and relevant in current times.

Specifically, automated aspect extraction models have been used to analyse online reviews [12]. What is meant by the automated extraction models is that such models can obtain information about the aspects present in a review. Think, for example, of a laptop review where a customer writes how the keyboard is nice and responsive. The model would be able to extract the aspect of “keyboard”. Then if the model is trained with sentiment analysis it is also able to discern that the review was positive on this aspect. This shows how these models are a great way of gaining insight into the positivity of a large number of reviews in a short amount of time. Unfortunately, creating automated models for aspect extraction is difficult, and developing them can cause developers to run into problems [21]. Finding out which problems occur and how to solve them is essential in the future improvements of aspect extraction models. To further the research in this field the following research question has been formulated.

What problems arise when predicting product specifications by performing automated aspect extraction on the product reviews?

With this research, the challenges that might be faced when doing automated aspect extraction were uncovered and hopefully, future researchers can benefit from the knowledge. This has been done by firstly annotating data of laptop reviews on Amazon, of which a lot of data has been collected. This annotated data was then used to train two kinds of models to predict aspects of these reviews. A supervised approach and an unsupervised approach. During the process of this research and after analysis of the results it has become clear that using a supervised approach is more suitable for this task. Even though the data annotation process is time-consuming, the difficulties and inaccuracies of the unsupervised model turned out to be a bigger problem. Hopefully, this research will help future researchers so that they can avoid these problems.

2. RELATED WORK

With the rise in the number of online reviews that now accompany most of the products you can find online, the research into aspect extraction and sentiment analysis has also increased over the years. The problem is that even after all this research it is still difficult to understand the workings of such a complicated model. The models are imperfect and sometimes make decisions that would not make sense to humans. This is a problem, because when models make mistakes that humans can not understand it drastically lowers the perceived accuracy of the model [19]. This is why further research is still needed.

Research on the topic of aspect extraction can be classified into four different approaches [2, 12]. There are supervised approaches with labelled data of which CNN [14] or LSTM [11, 22] are some well-known examples. Furthermore, there are semi-supervised models that only need some annotated data to work. These models make use of clustering of data and graph-based algorithms. And then unsupervised models do not require annotated data at all but are sometimes less accurate because of this [3, 13]. Then finally, there are rule-based models, which are not really artificial intelligence, since there is no learning process involved. They instead make use of a set of rules that the model follows [15].

Interesting and relevant research was done on a deep convolutional neural network (CNN) [14]. This is because the CNN models showed great potential for aspect extraction and thus could also be used as a model for this research. CNN shows a way of creating a non-linear supervised classifier. This is done by creating a deep neural network with multiple layers where each layer consists of visible and hidden neurons, where in each iteration the weights between these neurons are updated to optimize the outcome. Also, the hidden layer consists of Z groups of dimension $(Lx - nx + 1) * (Ly - ny + 1)$ where n is the filter kernel and L consists of the sentences and word embeddings.

Next to that, it is also important to know which features you want to look at from a customer perspective. Since this research will work with Amazon laptop reviews, it is good to know which properties of the laptop are of high value for the customers. Research on this topic has already been done and it came forward that, of course, the price is important and the core technical features are of value to the customer [1, 4]. Here you can think about having a good screen strength and a fast processor with a responsive keyboard[16]. Furthermore, the post-purchase services can also increase the satisfaction of the customer. These would be services like having regular updates and performing reparations if they are needed [4]. Because of this, those will be the main aspects of the product that this research will focus on and also the aspects that will be focused on for annotation.

3. METHOD

In general, the research will consist of annotating data using a dedicated annotation tool. The annotated data will be pre-processed so that it is ready to feed into the models. A supervised model will be trained and evaluated just as an unsupervised model. This will reveal the positives and negatives of both approaches and show which of the two methods works best.

3.1 Data Preparation

First of all, the dataset that will be worked with consists of a collection of Amazon laptop reviews. These reviews range from short 1 sentence reviews, to very detailed reviews that are a whole paragraph. These reviews are also spread over different kinds of laptops of different brands such as Lenovo, Mac, Dell, and HP.

Next to that, part of the research includes the annotating of data, since annotated data is useful for machine learning models. Unfortunately annotating data can be a quite time-consuming task, but an annotation tool will be used to speed up the process as much as possible. Then it is important what aspects will be annotated, because annotating everything would be too much. There will be a focus on the most important aspects. These will be the price, the processor, the keyboard, the screen, battery life, lifespan, and weight. Furthermore, the aspects will be annotated including sentiment analysis, which will either be negative, neutral, or positive. How this will work is when annotating a review you split it up into smaller sentences that are relevant to the aspect. These sentences will be labelled and given a sentimental value. It is important to note that the sentiment is not used in the models for sentiment analysis. It was purely part of the annotation process and can be used in future research.

Before models can be trained and used on data, the data must be represented and used in the best ways possible. So, before data is given to the model, the data is first stemmed and filtered [5]. The words in the data will be filtered for a list of stop words that do not give any useful information. Think of words like "the", "a", "to" and so on. Words that are not filtered will be reduced to their root form, which also helps with determining the frequency of these words in the documents.

3.2 Model Creation

After the data pruning, the automated models' training can start. For this research, two models will be created. One supervised model will use the annotated data and one unsupervised model will use the data without annotations. This way it is possible to document the differences and advantages of both types of models and see the different problems that can be run into. For the supervised model, LSTM was used, which was chosen for its natural language processing capabilities. Then for the unsupervised model, Fuzzy c-means clustering was used, because it allows for multiple aspects to be assigned to one data point. This is not possible with regular clustering algorithms such as k-means.

3.3 Analysis of Results

The final step of the research is to analyse the results of the models, which will be done in two ways. There will be a quantitative analysis with metrics. The most important metrics for aspect extraction models with sentiment analysis are precision, recall, and accuracy, which are calculated based on false and true positives and negatives where higher values indicate a better model.

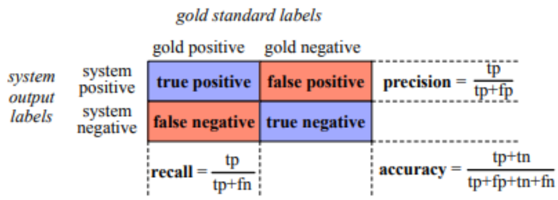

fig 1. Precision, recall, and accuracy [8]

To assess if a model works the way it is expected to work, it is also important to perform a qualitative analysis of the model. For aspect extraction, this means that the researcher will go over some of the aspects that were guessed wrongly and try to understand why the model made wrong predictions [7]. If, for example, the review was structured in an incorrect way that would make it difficult to read, even for a human, it would make sense that it guessed wrong and that would be acceptable. But, if the model makes inaccurate predictions on good reviews, that would be strange and might show that a model needs to be reworked. This way of reviewing a model is quite time-consuming, but it is very useful for understanding the workings of the model.

4. SUPERVISED MODEL

Before training a supervised model, it is important that there is enough annotated data for the model to learn from. The annotating of 1000 reviews can take up to 1,5-2 weeks, which goes to show how time-consuming this task can be. The data was annotated in the following way. Seven important aspects of laptops were chosen and these were annotated in the review. The aspects consist of price, processor, screen, keyboard, battery, lifespan, and weight.

After the data has been annotated it still can not be directly fed into a model, because it still needs to be pre-processed. For this research in particular, the full reviews had to be split up into singular sentences. This was done so that there would not be too many aspects in one data point. This would make it easier for the models to differentiate between aspects and understand the semantics behind them.

Furthermore, part of preprocessing the data was converting the reviews to lowercase, removing special characters and stop words. Also, the data was stemmed and tokenized. This has to be done because the special characters usually do not provide extra information. Next to that, it is important that “laptop”, “Laptop”, and “laptops” are all considered to be the same word with the same meaning for the model.

The model that was used in this research is an LSTM model, which stands for Long Short-Term Memory. The LSTM is a type of recurrent neural network. However, since it has the capability of maintaining information over a longer period of time due to its memory cell, it is great for understanding the connection between words throughout a sentence or document.

The specifics of the model used in the research are that the GloVe 6B of size 300 (see appendix) was used to help the model understand the semantic meanings of the words. Then the LSTM layer was added with 64 units, which is enough to understand the complexity of the sentences, but does not make the model too complex. Making the model too complex can lead to overfitting and choices that cannot be explained. This layer uses a tanh activation, which helps normalize the cell state between values of -1 and 1. Then finally, there is a dense layer with the size of the amount of aspects, which is 7 in this case. This layer uses sigmoid activation because this is necessary for multi-label classification. It is important to understand that SoftMax is also used for multi-class models, but SoftMax will assume that only one of the classes belongs to the data point. The sigmoid function gives all labels an independent probability, meaning multiple classes can belong to a single data point. Then the model is optimized using Adam and the loss function is binary cross-entropy. binary cross-entropy needs to be used instead of categorical cross-entropy, because the model has several outputs, since for each sentence it needs to evaluate for each aspect label whether it is in this sentence.

5. UNSUPERVISED MODEL

Then for the unsupervised model, the same reprocessed data as mentioned before has been used. A very popular unsupervised way of training on data is K-means. Here K amount of clusters are made and each datapoint is assigned to one of these clusters. However, since this research is working with sentences that can contain multiple aspects, it is not as simple as assigning the sentence to a single cluster.

To solve this a variation of K means was used, which is Fuzzy c-means. With Fuzzy c-means, it is possible to assign probabilities for each data point to belong to each cluster. This is because it outputs a membership matrix where the sum of the memberships for each cluster is 1. This way more complex sentences with multiple aspects can also be represented as a result.

For training the unsupervised model different text representations were used, similar to how an embedding was used for the supervised model.

The first text representation was made using a CountVectorizer to create a document term frequency matrix. This represents the data in a way where for each document (row) the amount of times a word (column) occurs in this document is stored.

However, in the context of text mining Term Frequency Inverse Document Frequency is often a better approach than a simple document term frequency. This is because, unlike regular term frequency, this approach takes into account that some words might occur often, but simply because they occur often in every document. This decreases the importance of the appearance of these words. For example, in laptop reviews, the word “laptop” might occur a lot, but it will not tell anything significant about our aspects.

Then there is an even more sophisticated method, which is the Sentence Transformer. In this case, the sentence transformer can be used to embed the documents into a dense vector. This way the dimensionality of the data is reduced and because the model is already pre-trained, it will give similar values to words that have similar meanings. This can be beneficial for training the model because it can help understand the relationship between words.

6. RESULTS

First of all, to gain more insight into how the aspects are distributed over the individual sentences, the total number of sentences that mention an aspect was plotted.

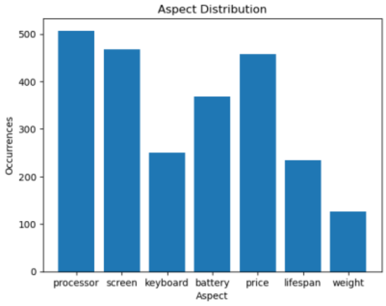

Fig 2. Distribution of Aspects over Reviews

In the image above you can see how certain aspects are talked about more often in laptop reviews than certain other aspects. Clearly, the price, processor, and screen are deemed to be the most important to customers.

Upon further examination, it also appeared that a little over 56% of the sentences in the dataset did not mention any of the aspects at all. 33% of the sentences mentioned a single aspect. 7% of the sentences mentioned exactly 2 aspects and thus the remaining 4% of the sentences mentioned 3 aspects or more. Not a single sentence had 6 or all 7 aspects in it. This makes sense because that would have to be a very big and crowded sentence.

An example of a sentence with three aspects would be:

[1 1 1 0 0 0 0]

"my summary is: you live with the screen every second you attend to a computer, the keyboard less, and everything else far less (trackpad, cpu, etc.)."

Here, the three 1's stand for processor, screen, and keyboard. These are set to 1 because they are mentioned in the sentence.

6.1 Supervised Model

At first, the results of the supervised model were quite disappointing. It got accuracies of around 30-40% which was not very impressive. However, this was the first model that was created without an embedding layer. Also, the training data that the model was fed were complete reviews instead of individual sentences. This likely made it more difficult for the model to learn from the data while predicting all the correct aspects in a review is also more difficult.

Another problem that occurred was that the model was acting as a binary decision model instead of deciding for each aspect whether it was present or not. What happened is that in the train set it could get a high accuracy, but on the test set, it got around 50%. This was because on the test set the model would only either predict [0,0,0,0,0,0,0] or [1,1,1,1,1,1,1], which is of course very wrong. This cause the model only to be able to recognize when the sentences talked about an aspect, but not the aspect it was talking about.

When the model was acting as a binary model it still got around 50% accuracy, but this was only because a lot of the data does not have any aspects in it. This means that accuracy gives a wrong view of the capacity of the model. As you can see in the image below, the precision, recall, and f1-score were way lower in this scenario.

```
Accuracy: 0.43829219479653103
Precision: 0.07735849056603773
Recall: 0.054959785522788206
F1 Score: 0.06426332288401254
```

Fig 3. Metrics for a binary model

Then, when adjusting the model so that it was looking at the aspects individually and decided for each one whether they were in the sentence, the accuracy stayed around the same level. This is also because with so many labels is it quite difficult to get all of them exactly right. However, doing this the precision, recall and f1 score increased significantly, which indicates that the model was working better than the previous model that had a "higher accuracy".

However, that was not the only change necessary to improve the f1 score of the model. Since the model had some difficulties deciding when none of the aspects were present, an extra label was added to the aspects that would only be set to 1 if all other aspects were not present. When evaluating the predictions of the model they seemed a lot more logical and when mistaken it was usually only 1 aspect that the model was either missing or labeled unnecessarily.

```
Accuracy: 0.4408817635270541
Precision: 0.5301204819277109
Recall: 0.4835164835164835
F1 Score: 0.5057471264367817
```

Fig 4. Metrics for a multi-label model

Furthermore, while tuning the hyperparameters of the model, the problem of an exploding gradient occurred from time to time [20]. This happens when values in the loss function become too large and keep becoming larger to the point where the model is unable to learn any further. The problem occurred when using too many layers or units in the model.

Then, after reformatting the data, and adding an embedding layer to the model its performance started improving. This, together with the finetuning of some hyperparameters, such as the number of units, the loss and activation function, and the optimizer, the model performed a lot better with an accuracy above 95% on the train set around 70% on the validation set. However, testing it on some new data outside of the dataset rendered results of a little over 45%, which could point to overfitting of the model.

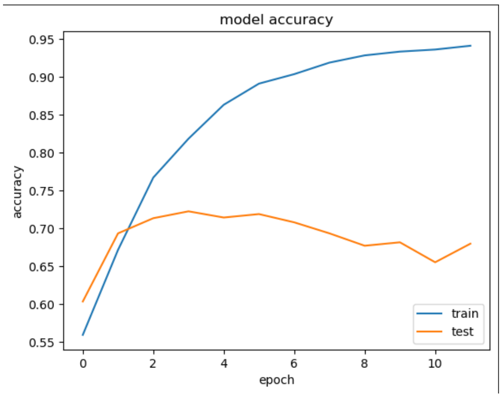

fig 2. Predicted Accuracy of LSTM Train vs Test

6.2 Unsupervised Model

The results for the unsupervised model were a bit underwhelming. This is because the first model that was made was made using K-means. Since this model does not take into account that one data point can belong to multiple clusters it did not give the intended results.

Switching over to Fuzzy c-means seemed to be the solution for this, but unfortunately, after training the model on DTF and TF-IDF, both resulted in all the clusters being almost at the exact same point and the model thinking that each data point belonged to each cluster. Changing the number of clusters did not change this behaviour in any way. Even when checking for each data point what the cluster with the biggest likelihood was, the model clearly did not understand what the aspects were. Which was to be expected with the clusters being so close together.

In a final attempt to get the cluster centres further away from each other, the data was encoded using a pre-trained sentence transformer. This transformer created embeddings for the data points with the hopes of allowing the model to understand the relationship between words better. At first, it seemed to have somewhat of an impact, because the cluster centres were ever so slightly farther away from each other now. Unfortunately, when checking the individual data points and their closest cluster it appeared that again there was no coherency in the model predictions, and still, the clusters were too close to each other.

7. DISCUSSION

Firstly, the LTSM model had some nice results, but could still be improved. The advantage it has of being able to store information and understand the concepts of longer sentences showed [11]. Next to that, with the current structure of the model training, it was not time-consuming. However, the dataset that it was trained on is not that large. With more annotated data there is a possibility of the model improving. This is because LSTM models tend to become more stable when they use more data [22]. Next to that, to reduce overfitting and make the model robust it might be wise to randomize the data. Right now the model is trained on reviews for specific laptop reviews, while the test set consists of reviews of other laptops. Mixing these could improve the model's versatility. Considering the results and ease of making the model, the time spent annotating data seems worthwhile.

Then, for the unsupervised model, the results were worse than hoped. Again, maybe increasing the dataset could help the model improve. In the case of a supervised model getting more data is easier, so it is worth a try. However, since there is no labeled data it is very hard to train an unsupervised model for aspect extraction. This is because it is more difficult to guide the model in the right direction. Next to that, it is hard to see why this model makes certain choices at times when it does not make the right choices.

Also, different embeddings might make a difference, since switching to the sentence transformer seemed to have some impact. Furthermore, extra data preprocessing could be done to guide the model in the right direction. More words that we know will not help identify the aspects that can be removed so that only keywords remain. However, doing this would defeat the purpose of an unsupervised model, since you are then specifically going through the data. In that case, it seems easier to just annotate the data and then use it in a supervised model, since that will certainly work.

It does not seem useful to use both these approaches in parallel. It would make more sense to decide before you start on a project whether it is possible to get access to annotated data. Based on this, a choice can be made on either using a supervised approach or an unsupervised approach. Making both of them at the same time does not have any added value. Furthermore, using the supervised model is more feasible for being used in production, because it gives a bigger guarantee for results. Next to that, it is easier to modify to make it work correctly on the data that is given.

Then finally, to get back to the research question of what problems arise when predicting product specifications by performing automated aspect extraction on the product reviews? As this research has shown the main problem for these models is to get them to understand the semantics of the reviews. This has proven to be the most difficult for an unsupervised approach.

8. CONCLUSION

To conclude, several problems arise when performing automated aspect extraction on product reviews. This already started in the data annotation process. It is a very time-consuming task and it is difficult to keep consistent over a large number of reviews. However, the preprocessing was pretty straightforward. The only important problem that arose is that leaving the reviews as they are does not work as well as splitting them up into separate sentences. This way you might lose some deeper context, but it is still worth it.

Then, after annotation and preprocessing were done, the creation of the supervised model was quite doable. The model quickly started to understand the aspects that occurred in the data and by making small adjustments the accuracy got better and better. In the end, it is worth the time it takes to annotate the data.

Especially when comparing the supervised model to the unsupervised model it becomes clear that supervised models simply work better for aspect extraction in laptop reviews. The time that is saved skipping the annotation will go into trying to understand and modify the unsupervised model anyway. And even then, it is not certain that this will work. Especially with laptop reviews where terms and phrases might sometimes overlap for different aspects.

However, more research can still be done on how unsupervised models help in aspect extraction in the future. Especially if there is access to more data.

9. REFERENCES

[1] Andrea Papenmeier, Alexander Frummet, and Dagmar Kern. 2022. "Mhm..." – Conversational Strategies For Product Search Assistants. In Proceedings of the 2022 Conference on Human Information Interaction and Retrieval (CHIIR '22). Association for Computing Machinery, New York, NY, USA, 36–46. https://doi-org.ezproxy2.utwente.nl/10.1145/3498366.3505809

[2] Davoodi, L. (2023). ENHANCING THE UNDERSTANDING OF E-COMMERCE REVIEWS THROUGH ASPECT EXTRACTION TECHNIQUES: A BERT-BASED APPROACH. 36th Bled eConference Digital Economy and Society: The Balancing Act for Digital Innovation in Times of Instability, 233.

[3] Eisenstein, J. (2019). Introduction to Natural Language Processing. MIT Press. gt-nlp-class/notes/eisenstein-nlp-notes-10-15-2018.pdf at master · jacobeisenstein/gt-nlp-class · GitHub

[4] Factors influencing consumers' laptop purchases dr. V. Aslıhan Nasır, Sema Yoruker, Figen Güneş and Yeliz Ozdemir Bogazici University, Istanbul, Turkey abstract. (z.d.). https://ininet.org/factors-influencing-consumers-laptop-purchases-dr-v-aslhan-nas.html

[5] Gupta, D., & Ekbal, A. (2014, August). IITP: supervised machine learning for aspect based sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 319-323). https://aclanthology.org/S14-2053.pdf

*[6] Hu, N., Liu, L. & Zhang, J.J. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. Inf Technol Manage **9**, 201–214 (2008). https://doi-org.ezproxy2.utwente.nl/10.1007/s10799-008-0041-2*

[7] Jeon, S.W., Lee, H.J., Lee, H., Cho, S. (2019). Graph Based Aspect Extraction and Rating Classification of Customer Review Data. In: Li, G., Yang, J., Gama, J., Natwichai, J., Tong, Y. (eds) Database Systems for Advanced Applications. DASFAA 2019. Lecture Notes in Computer Science(), vol 11448. Springer, Cham. https://doi-org.ezproxy2.utwente.nl/10.1007/978-3-030-18590-9_13

[8] Jurafsky, D., & Martin, J. H. (2014). Speech and Language Processing. Pearson Speech and Language Processing (stanford.edu)

[9] Jyoti Prakash Singh, Seda Irani, Nripendra P. Rana, Yogesh K. Dwivedi, Sunil Saumya, Pradeep Kumar Roy, Predicting the "helpfulness" of online consumer reviews, Journal of Business Research, Volume 70, 2017, Pages 346-355, ISSN 0148-2963, https://doi.org/10.1016/j.jbusres.2016.08.008.

[10] Lackermair, G., Kailer, D., & Kanmaz, K. (2013). Importance of online product reviews from a consumer's perspective. Advances in economics and business, 1(1), 1-5. https://www.intotheminds.com/blog/app/uploads/importance-product-reviews-lackermair.pdf

[11] Londhe, A., & Rao, P. P. (2021). Aspect based sentiment analysis–an incremental model learning approach using LSTM-RNN. In Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021, Nashik, India, April 23–24, 2021, Revised Selected Papers, Part I 5 (pp. 677-689). Springer International Publishing.

[12] Maitama, J. Z., Idris, N., Abdi, A., Shuib, L., & Fauzi, R. (2020). A Systematic Review on Implicit and Explicit Aspect Extraction in Sentiment Analysis. IEEE Access, 8, 194166-194191. https://doi.org/10.1109/access.2020.3031217

[13] Mauro Dragoni, Marco Federici, Andi Rexha, An unsupervised aspect extraction strategy for monitoring real-time reviews stream, Information Processing & Management, Volume 56, Issue 3, 2019, Pages 1103-1118, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2018.04.010.

[14] Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems, 108, 42-49. https://doi.org/10.1016/j.knosys.2016.06.009

[15] Poria, S., Cambria, E., Ku, L. W., Gui, C., & Gelbukh, A. (2014, August). A rule-based approach to aspect extraction from product reviews. In Proceedings of the second workshop on natural language processing for social media (SocialNLP) (pp. 28-37).

[16] Rachdian, A. O., Suryadi, D., & Fransiscus, H. (2022). Identification of Customer Needs from Product Reviews using Topic Modeling and Aspect-based Sentiment Analysis. International Journal Of Computing And Digital System/International Journal Of Computing And Digital Systems, 12(6), 1383–1394. https://doi.org/10.12785/ijcds/1201111

[17] Zhao, Y., Wen, L., Feng, X., Li, R., & Lin, X. (2020). How managerial responses to online reviews affect customer

satisfaction: An empirical study based on additional reviews. Journal of Retailing and Consumer Services, 57, 102205. [*https://doi.org/10.1016/j.jretconser.2020.102205*](https://doi.org/10.1016/j.jretconser.2020.102205)

[18] Elwalda, A., & Lu, K. (2016). The impact of online customer reviews (OCRs) on customers' purchase decisions: An exploration of the main dimensions of OCRs. Journal of Customer Behaviour, 15(2), 123-152. https://doi.org/10.1362/147539216x14594362873695

[19] Andrea Papenmeier, Dagmar Kern, Daniel Hienert, Yvonne Kammerer, and Christin Seifert. 2022. How Accurate Does It Feel? – Human Perception of Different Types of Classification Mistakes. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 180, 1–13. [*https://doi-org.ezproxy2.utwente.nl/10.1145/3491102.3501915*](https://doi-org.ezproxy2.utwente.nl/10.1145/3491102.3501915)

[20] Rehmer, A., & Kroll, A. (2020). On the vanishing and exploding gradient problem in Gated Recurrent Units. IFAC-PapersOnLine, 53(2), 1243-1248. [*https://doi.org/10.1016/j.ifacol.2020.12.1342*](https://doi.org/10.1016/j.ifacol.2020.12.1342)

[21] Tubishat, M., Idris, N., & Abushariah, M. A. (2018). Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges. Information Processing &Amp; Management, 54(4), 545-563. [*https://doi.org/10.1016/j.ipm.2018.03.008*](https://doi.org/10.1016/j.ipm.2018.03.008)

[22] Peng, A., Zhang, X., Xu, W., & Tian, Y. (2022). Effects of Training Data on the Learning Performance of LSTM Network for Runoff Simulation. Water Resources Management, 36(7), 2381-2394. [*https://doi.org/10.1007/s11269-022-03148-7*](https://doi.org/10.1007/s11269-022-03148-7)

10. APPENDIX

During this research, the author used ChatGPT to aid in the process of writing code. This usually occurs when receiving error messages where the tool could be used to give ideas for a solution. After using this tool/service, the author reviewed and edited the content as needed, because ChatGPT will not simply give the perfect answer and the code still needs to be modified so the author takes full responsibility for the content of the work."

The GloVe 6b embedding can be found here: https://nlp.stanford.edu/projects/glove/