Mining Association Rules in Commuter Feedback Comments from Facebook of Swiss National Railways (SBB) using Apriori Algorithm

Patrick Blatter[1] and Farshideh Einsele[2]

[1] Distance University of Applied Sciences, Zollstrasse 17, 8005 Zurich, Switzerland
[2] Bern University of Applied Sciences, Brückenstrasse 63, 3002 Bern, Switzerland
farshideh.einsele@bfh.ch

Abstract. Nowadays, all kinds of service-based organizations open online feedback possibilities for customers to share their opinion. Swiss National Railways (SBB) uses Facebook to collect commuters' feedback and opinions. These customer feedbacks are highly valuable to make public transportation option more robust and gain trust of the customer. The objective of this study was to find interesting association rules about SBB's commuters pain points. We extracted the publicly available FB visitor comments and applied manual text mining by building categories and subcategories on the extracted data. We then applied Apriori algorithm and built multiple frequent item sets satisfying the minsup criteria. Interesting association rules were found. These rules have shown that late trains during rush hours, deleted but not replaced connections on the timetable due to SBB's timetable optimization, inflexibility of fines due to unsuccessful ticket purchase, led to highly customer discontent. Additionally, a considerable amount of dissatisfaction was related to the policy of SBB during the initial lockdown of the Covid-19 pandemic. Commuters were often complaining about lack of efficient and effective measurements from SBB when other passengers were not following Covid-19 rules like public distancing and were not wearing protective masks. Such rules are extremely useful for SBB to better adjust its service and to be better prepared by future pandemics.

Keywords: Opinion Mining, Data Mining, Association Analysis, Apriori Algorithm, Online Customer Feedback Mining, Text Mining

1 Introduction

Swiss Federal Railways (SBB) is the national railway company of Switzerland. SBB was founded in 1902 as a government institution, but since 1999 Swiss cantons are participating in its ownership as well. While SBB is ranked first among national European rail systems in 2017 due to its intensity of use, quality, and reliability, it is also suffering from breakdowns and security problems, struggle to keep timetable and overcrowded trains. At the same time the wages in the SBB boardroom have been highly risen. (Handelszeitung, 2019). Additionally, the company suffers from outdated rolling stock as well as its dependency on foreign rolling stock.

Like many other companies SBB is on Facebook and has a visitor page, where the commuter can leave their comments. These comments are written in majority in German (approx. 80%) but also French and Italian. Switzerland is a multi-language country, and these are its official written and spoken languages. The idea of this study was to mine association rules in commuter feedback and gain an understanding of their opinion and satisfaction rate in various categories.

Consumer reviews are a reliable source of market response and text mining on these unstructured textual review data which can provide significant insight about consumers opinion. That is why many researchers have used text mining methods in different business contexts to extract meaningful insights and applied association mining to extract association rules to summarize customer opinions. Mining association rules as a method for customer opinion mining has been used by various researchers. The main objective of association mining is to determine rules that are extracted from a specific database by a pre-determined minimum percentage of support and confidence. (Wong, 2009) extracts feature and opinion using part-of-speech tagging on each review sentence and applies a-priori algorithm to gain association rules to summarize customer review opinions. Liang et al. (Liang, 2017) used a set of trivial lexical-part-of-speech patterns to find the optimal number of topics from online feedbacks of Chinese users along with Apriori algorithm and found out strong rules in each topic. Abulei (Abuleil, 2017) has utilized NLP techniques to generate some rules that helps to understand customer opinions and reviews written in the Arabic language. Peska et al., (Peska, 2017) uses a specific user behavior pattern as implicit feedback, forming binary relations between objects i,e. analyzing rule relations though it is not insignificant to capture the implicit relations and even to make the actionable well-known rules. (Hilage, 2011) describes using association rule mining, rule induction technique and Apriori algorithm to understand the correct buying behavior of the customer. (Huang, 2011) use the relevance feedback mechanism in the collaborative recommendation system and adopt statistical attribute distance method to calculate the customer feedback correlations. They then apply multi-tier granule mining to find association rules to provide customers more relevance information.

Knowledge discovery in Databases (KDD) has been proposed by Fayyad et al (Fayyad, 1996) as a framework to gain relevant knowledge in large and often unstructured databases. The process of knowledge discovery contains 6 steps to obtain knowledge from one or more databases. The process starts with retrieving data from various sources, clean and pre-process them and integrate them into one single database. apply data mining and finally discover novel and useful knowledge.

In our study, we apply KDD steps. Since in our case data is available in form of textual comments from commuters, we use some manual text mining techniques to gain categories and then build a relational database with each category as a table containing several sub-categories. In data mining step we have chosen Apriori algorithm over FP-growth, due to its simplicity, since our data set was relatively small and Apriori algorithm has proven to be as effective as FP Growth. We have gained interesting association rules that summarize the customer's opinion of SBB for the entire year of 2020. 2020 was a special year driven by Corona pandemic, hence a substantial number of comments were related to Covid-19 pandemic which were combined with different other categories extracted through text mining.

2 Data Selection

Data was extracted using an online extraction tool as a headless browser that programmatically visits a website and follows a sequence of predefined interactions. A linguistic recognition of the individual feedbacks was carried out using Open Refine (Open Refine, 2020) to remove all non-German-speaking feedback

3 Pre-Processing

The data was pre-processed with the “Open Refine” tool. Open Refine is a software provided by Google freely available and a part of CS&S (Code for Science & Society, 2021)). The messages were cleaned up to the point that all line breaks (tag: \ n) and emojis were removed. See Fig. 1 (removing line break tag) & Fig. 2 (removing emojis)

Expression

```
return value.replace("\n", "");
```

Fig. 1. Removing line break tags

Expression

```
import re

def remove_emojis(data):
    emoj = re.compile("["
        u"\U0001F600-\U0001F64F"  # emoticons
        u"\U0001F300-\U0001F5FF"  # symbols & pictographs
        u"\U0001F680-\U0001F6FF"  # transport & map symbols
        u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
        u"\U00002500-\U00002BEF"  # chinese char
        u"\U00002702-\U000027B0"
        u"\U00002702-\U000027B0"
        u"\U000024C2-\U0001F251"
        u"\U0001f926-\U0001f937"
        u"\U00010000-\U0010ffff"
        u"\u2640-\u2642"
        u"\u2600-\u2B55"
        u"\u200d"
        u"\u23cf"
        u"\u23e9"
        u"\u231a"
        u"\ufe0f"  # dingbats
        u"\u3030"
                      "]+", re.UNICODE)
    return re.sub(emoj, '', data)

return remove_emojis(value)
```

Fig. 2. Removing emojis

Apart from the above removals, further text mining steps were not carried out because the textual data was not automatically processed. Instead, we divided the feedbacks into categories and sub-categories in order not to lose the context of feedbacks. At the same time, an attempt was made to keep the number of categories & their subcategories as small as possible to avoid a kind of overfitting, which in turn can lead to many non-meaningful association rules. Each feedback is always assigned to at least the overall-impression category. This category records the state of mind of the person who wrote the message. In addition, each feedback is divided into one of the following time groups based on the publication date of the feedback: night (24:00 - 06:00), rush hour morning (06:00 - 09:00), marginal time (09:00 - 11:00 and 2 p.m. - 4 p.m.), noon (11 a.m. - 2 p.m.), rush hour evening (4 p.m. - 7 p.m.) and evening (7 p.m. - midnight), this division was performed automated. Fig. 3 shows the categories, their sub-categories and number of corresponding feedbacks in percentage. This type of classification is also called {multi-label classification in the machine learning environment. The aim is to assign text fragments such as sentences, sections, or even entire documents to one or more labels (paperswitchcode, 2021)

Table 1. Categories & Sub-Categories

Category	**Sub-Category**	**Number of Feedbacks %**
Gesamteindruck	Unzufrieden, Neutral, Zufrieden	100.0
Zeitgruppe	Stosszeit Morgen, Mittag, Randzeit, Stosszeit Abend, Abend	100.0
Service	Unternehmen, Kundenservice, Information, Ansage, Frage, Sonstiges, Auskunft, Kritik, Lob, Mitarbeiter	21.03
Corona	Keine Maske, Prävention, Frage, Sonstiges, Fahrplan, Maske, Ticket	19.80
Tickets	Zurücklegen/Rückerstattung, Unverständliche Busse, Preis zu hoch/Preisgestaltung, Kauf nicht möglich, Frage, Sonstiges, GA, Studenten-GA, SwissPass, Sparbillet, Klassenwechsel, Hundepass	17.54
Infrastruktur	Zug defekt, Verbesserungsvorschlag, Temperatur, Zug verkürzt, Behindertengerecht, Frage, Sonstiges, Sauberkeit, WC, Stauplatz, Tür automatisch	15.53

Zuverlässigkeit	Verspätung, Zu früh, Ausfall, Anschluss verpasst, Frage, Sonstiges, Störung	13.61
Sonstiges	Frage zu Baustelle, Diebstahl, Bettler, Frage, Sonstiges	12.4
Fahrplan	Fehlerhaft, Verbesserungsvorschlag, Frage, Sonstiges, Einhaltung, Streichung	7.84
Kapazität	Überfüllt	7.60
App	Störung, Fehlerhaftes Angebot, Frage, Sonstiges, Fahrplan	5.76
Bahnhof	Beschädigt,Vandale, Frage, Sonstiges, Sauberkeit, Infrastruktur	4.54
Mitreisende	Falsches benehmen, Frage, Sonstiges, Korrektes benehmen	4.36
Restaurant	Angebot, Geschlossen, Frage, Sonstiges, Kein Speisewagen, Reservation	1.74

Multiple researcher report of using automatic categorization in their works. Liu et al (http://arxiv.org/abs/2008.06695) achieved a Micro F1 score of a maximum of 72% in their proposed procedure in a data set consisting of 55,840 samples with a total of 54 possible labels to choose from (Liu, 2020). The Micro F1 score describes the quality (correct assignment of artifact and label) of the model used. Ibrohim and Budi achieved an accuracy of 66% in the classification of 5561 tweets that could be assigned to six possible labels (three main categories, three sub-categories) (Ibrohim, 2019). Based on the results of their finding, we decided to take a manual approach because a correct prediction of labels in the range of 66 – 72% is not satisfying given the small number of data set

Furthermore, based on the already small total amount of the database (1000 Feedbacks for 2020), we decided that an accuracy of less than 85% cannot be regarded as sufficient due to the falsified results. In addition to this, the poor quality of the feedback itself speaks in favor of manual classification. Spelling errors, Helvetisms (a.k.a Swiss German spoken dialect) and sarcasm as well as the mixing of German and Swiss German were all supplement factors that made it difficult to assign correct categories automatically.

To sum up, to achieve the best possible results from the association analysis, an automated process was therefore withdrawn. The Categories & sub-Categories are therefore intuitive. This means that the classified data can turn out differently depending on the person who performs this task. Fig. 3 shows the categories, the sub-categories, and the number of corresponding feedbacks in percentage.

4 Data Transformation

Data has been transformed into a database using a dimensional scheme with the category overall-impression as the fact. The other categories are viewed as dimensions of feedback, as they describe the context of feedback in more detail. These tables contain the associated subcategories as attributes in a binary form (0 = feedback does not belong to this category, 1 = feedback belongs to this category). The fact table represents the feedback in the center of the scheme and uses a binary form to describe the affiliation to the other categories. The fact table includes all main categories as non-optional foreign keys to ensure that every feedback must explicitly describe whether it belongs to a category or not. As a result, the consistency and integrity of the data can be guaranteed as soon as feedback is inserted into the table scheme).

The tables are filled automatically using Python. The categorized data in the original CSV format served as the starting point. For this purpose, the file was read, and the python code was carried out, which fills the DB with the data from the file.

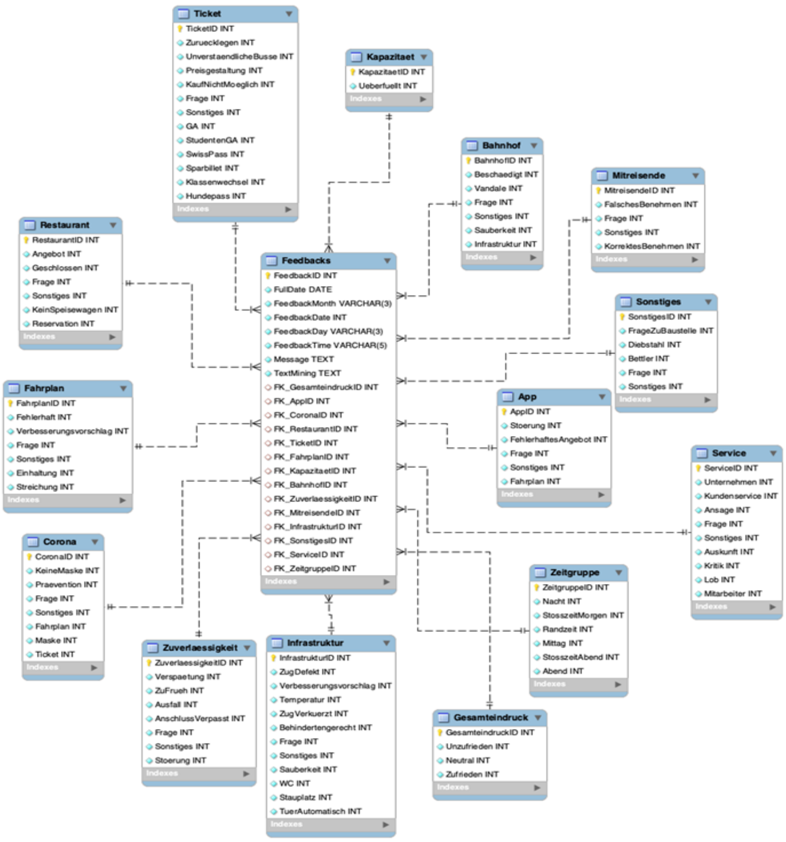

Fig. 3. Dimensional Database Scheme of Feedbacks

5 Data Mining

Due to our relatively small number of feedbacks (1000 feedbacks for 2020) and a relatively small number of assigned categories and sub-categories of the feedback, there is little difference in performance whether to use the a-priori or FP growth algorithm to determine the frequency sets. The rules resulting from the association analysis can in this case be achieved with both algorithms without being able to determine qualitative differences. On this basis, the decision was made to determine the frequent item sets using the a-priori algorithm.

5.1 Building frequent item sets

We have determined the possible combinations of the sub-categories within each category to build the frequent 2,3,4 item sets. As an example of this, if selecting the categories to build a frequent 3-item set were overall impression and service, then a possible frequent itemset could consist of the s satisfied, employee, praise. Figure 6 shows an overview of the distribution of the categories. It can be seen that numerous combinations are possible with the overall-impression category due to the frequency of this category. At the same time, further observations of the categories reliability and restaurant showed that there was no feedback relating to both categories. Furthermore, it was decided not to consider the category miscellaneous for further analysis, since no meaningful rules could be formed with these categories due to lack of related data. Furthermore, the greater the number of categories for determining the frequent item sets, the fewer possible meaningful combinations of categories there were. This is because there is only a small amount of feedback that includes four or more categories. For example, there is only one feedback in the database that has five categories. For this reason, only sets up to and including a maximum of four frequent item-sets were created. Fig. 6 shows category combinations from which frequent 2- and 3-item sets were created.

Table 2. Frequent item-sets

Frequent 2-itemsets	**Frequent 3-itemsets**
Gesamteindruck, Zeitgruppe	Gesamteindruck, Zuverlässigkeit, Infrastruktur
Gesamteindruck, Zuverlässigkeit	Gesamteindruck, Corona, Kapazität
Gesamteindruck, Corona	Gesamteindruck, Corona, Service
Gesamteindruck, Ticket	Gesamteindruck, Corona, Infrastruktur
Gesamteindruck, App	Gesamteindruck, Corona, Mitreisende

Gesamteindruck, Service	Gesamteindruck, Ticket, Service
Gesamteindruck, Infrastruktur	Gesamteindruck, Zuverlässigkeit, Service
Gesamteindruck, Restaurant	Zeitgruppe, Zuverlässigkeit, Infrastruktur
Gesamteindruck Bahnhof	Gesamteindruck, Zuverlässigkeit, Zeit-gruppe
Gesamteindruck, Fahrplan	Gesamteindruck, Ticket, Corona
Ticket, Service	Gesamteindruck, Service
Corona, Mitreisende	Gesamteindruck, Corona
Infrastruktur, Kapazität	Service, Zuverlässigkeit
Corona, Ticket	Gesamteindruck, Corona, Fahrplan
Zuverlässigkeit, Service	--

5.2 Building association rules

We have built association rules using python as follows: the records (feedbacks) of an itemset combination that are in the database were first loaded into the main memory with a read query. For this purpose, all possible columns in the database were determined for each item set that occurs (without a primary key, as this does not contain any useful information). As soon as these were available, a SQL statement was generated, which linked the column n with a table prefix until all columns to be read appeared in the statement. The statement is completed so that all categories of an item set are linked to the Feedbacks table by means of an inner join. The following parameters were used to create the frequent item sets and the association rules: For the generation of frequent item-sets there was a minimum support given that had to be achieved for it to be considered in the association rule mining process. A parameter defining how many items can occur in total within an association rule was passed along while building up to 4-itemsets and starting from 2. For every found association rule it was further checked if they meet a minimum confidence of 0.7.

6 Results (Association Rules)

We have built a total of 315 association rules, i.e., 75/2-itemsets, 197/3-itemsets and 43/4-itemsets association rules from our 2,3,4 frequent item-sets. Below are some examples of the gained rules:

6.1 Overall impression & Corona

A total of 227 feedbacks belonged to this item set. {Used minimum support: 0.1

1. In 86% of the cases, a missing mask also leads to dissatisfaction. Both occur in 3.75% of all feedback. There is a correlating dependency between premise and conclusion (lift = 1.17)
2. In 84.7% of the cases, the (lack of) prevention leads to dissatisfaction. This occurs in 10.21% of the total amount. There is a correlating dependency between premise and conclusion (lift = 1.15)

6.2 Overall impression & ticket

A total of 201 feedbacks belonged to this item set. Minimum support used: 0.1

3. In 96.0% of the cases, the pricing leads to dissatisfaction. This appears in 4.42% of all feedbacks. There is a positive correlation between premise and conclusion (lift = 1.62)
4. If for some reason the purchase of a ticket is not possible, the customer is dissatisfied with the experience in 84.0% of the cases. Both occur in 2.18% of the total amount and there is a positive correlation between premise and conclusion (lift = 1.49)

6.3 Overall impression, reliability, time

A total of 156 feedbacks belonged to this item set. Minimum support used: 0.1

5. In 82.6% of the cases, dissatisfaction at rush hour can be attributed to being late. All occur in 1.66% of the total amount. There is a positive correlation between premise and conclusion (lift = 1.18)
6. A delay in the evening leads to dissatisfaction in all cases. Together, these appear in 1.4% of all feedbacks. There is almost no dependency between premise and conclusion (lift = 1.01)
7. A delay in the morning rush hour leads to dissatisfaction in all cases. Together, these occur in 1.66% of all feedback. There is almost no dependency between premise and conclusion (lift = 1.01)
8. A delay at off-peak times leads to dissatisfaction in all cases. Together, these occur in 1.66% of all feedback. There is almost no dependency between premise and conclusion (lift = 1.01)

7 Knowledge Interpretation

In the following paragraphs, we give a brief interpretation of some of the interesting rules found in our data-mined categories. An overall interpretation was that especially dissatisfied customers left their feedbacks. Rules that contained the item Satisfied very often have not even reached the set minsup and were therefore ignored when generating the rules. In this case, the minimum support should have been set very low, which in turn would have caused the problem of generating too many rules.

7.1 Reliability

As a state-owned company, the main task of SBB is to transport commuters. Hence the commuters were especially dissatisfied when train was not on time or another expectation from their point of view could not be met. Rules found from the areas of reliability and time indicate that traffic is at its peak and is often delayed during rush hours in the morning as well as evening, whereas trains were mostly half full in the remaining hours. SBB should come with a better plan to answer this shortcoming.

7.2 Covid -19

A novel source of dissatisfaction has emerged from the initial lockdown situation in 2020 when commuters didn’t wear a protective mask and not complying with Distance rules result from overcrowded trains to prevent the Covid-19 virus. These rules turned out to be very interesting. Through increased train controls by SBB in combination with covid-related fines for commuters who do not wore the protective mask, SBB could have counteracted at least the point regarding non-compliance with the mask requirement. Part of the blame, however, lays in the decision of SBB to shorten train formations. This decision may seem a cost-optimizer and it may be justifiable since the offer was limited, however the actual demand from commuters was greater than anticipated.

In addition, the exemplary behavior of the SBB employees highlighted in relation to the prevention of the Covid-19 virus. Their Action has often led to explicit praise. Another positively correlating rule is dissatisfaction with cleanliness and Hygiene and the resulting inadequate prevention of Covid-19. Commuters wished, among other things, that the train doors would open by themselves, so that they come so little as possible into contact with material from potentially infected persons. SBB could improve the unsatisfied customers with a more transparent communication towards the passengers, to what extent cleaning was carried out.

Lastly, customers who have a relatively expensive yearly ticket (GA travelcard) were often dissatisfied, since it was not possible to receive compensation such as an extension of their subscription due to the reduced SBB offer. With a clear communication policy, it would have been possible to explain to customers even better why it was not possible to compensate.

7.3 Train ticket

The commuters wished for more goodwill by fines. Many fines were considered incomprehensible from the perspective of the commuters. This led to dissatisfaction and criticism in equal measure. For example, when for uncertain reasons the ticket purchase was not possible, the commuters were dissatisfied, since the commuters had to carry the risk of receiving fines or be delayed.

7.4 SBB APP

The App timetable and the ticket purchase with wrong prices were sources of discontent. Both points have turned out to be interesting rules. In some cases, delays or train cancellations were not apparent. In the worst case, there were even false reports of train connections. It also happened that, for example, certain tickets could be purchased at the wrong price. A fault within the app, on the other hand, which receives most of the feedback within this Category, show almost no dependency, which is why it cannot be said that a disorder leads to dissatisfaction.

7.5 Timetable

The deletion of a connection led often to dissatisfaction with those directly affected. This can be attributed to the fact that these individuals rely on exactly this connection and are now looking for alternatives as a result. Regarding a timetable optimization on the part of the SBB, it is obvious that SBB cannot meet all passenger needs. In this regard, only early information in connection with any reasons for the decision, should be put on place by SBB to minimize this dissatisfaction as much as possible and, if necessary, to make the appeal understandable to the commuters.

8 Conclusion & future work

The presented study reports of mining interesting association rules from Facebook-posted commuter feedbacks of Swiss National Railways (SBB). This study is the first public study on commuter feedback of SBB. We have used text mining and datamining specifically association analysis to by applying Apriori algorithm on the data. We have built different categories and built association rules within their different subcategories. Since the data set was very unbalanced, i.e. there were many more dissatisfied entries than satisfied ones, most of the rules were based on factors related to a dissatisfaction. Because of that rules that relate to satisfaction can be valuable as well. However, this is a known phenomenon by customer feedback on a company’s online page. This kind of feedback is fundamentally different from the posts where customers write their opinion for other customers like in TripAdvisor and share their good and bad experiences. Notwithstanding, the rules related to passenger’s discontent are not any less useful, as they provide insights into the customer's point of view and define specific pain points.

Most unsatisfied comments were found by late train during rush hours, deleted but not replaced connections on the timetable due to SBB’s timetable optimization, inflexibility of fines due to unsuccessful ticket purchase. Additionally, a considerable amount of discontent was related to the policy of SBB during the initial lockdown of the Covid-19 pandemic. Commuters were often complaining about lack of efficient and effective measurements from SBB when other passengers were not following Covid-19 rules like public distancing and were not wearing protective masks. Such rules of discontent give a valuable insight into what exactly needs to be done to meet the commuter's expectations.

During the implementation it was shown that some steps of text mining should be performed manually to categorize the data. In practice, this approach does not represent a scalable and efficient method in large data collections, since the process execution is time consuming. For our future work, we intend to perform automatic text categorization while gathering more data.

Another weakness of Apriori algorithm is the definition of a minimum support. It is important to find a mediocrity in order not to have to generate too many uninteresting rules and at the same time not to delete too many potentially interesting rules. This value can only be found by trial and testing.

To be able to address the problem of scalability present in this study, an increase in the accuracy of multi-label classification models would be desirable. This depends on various factors such as the number of labels, the amount of database and their balance. An experiment could be used to research whether a certain degree of accuracy with several labels can be achieved. Among other things, this would be It is also interesting to see which number of labels by the size of the data would lead to the predefined maximum accuracy (between 80%-100%) . We intend to consider the above-mentioned improvements in our future work.

References

Handelszeitung, https://www.handelszeitung.ch/unternehmen/si-kemme-am-saggsi-und-geen-am-zwai

Wong: A Method for Opinion Mining of Product Reviews using Association Rules, ICIS '09: Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, November 2009, Pages 270–274, https://doi.org/10.1145/1655925.1655973

Liang, , Mining product problems from online feedback of Chinese users, Emerald Publishing Limited 2017, https://doi.org/10.1108/K-03-2016-0048

T. Velmurugan, Mining Implicit and Explicit Rule for Customer Data Using_Natural Language Processing and Apriori Algorithm, January 2020International Journal of Advanced Science and Technology 29(9):3155 - 3167

Saleem Abuleil and Khalid Alsamara, "Using Nlp Approach For Analyzing Customer Reviews", pp. 117– 124, 2017. © Cs & It-Cscp 2017 Doi: 10.5121/Csit.2017.70112.

L. Peska and P. Vojtas, "Using implicit preference relations to improve recommender systems," Journal on Data Semantics, Vol. 6, No. 1, pp. 15–30, 2017.Hilage Tejaswini A, Kulkarni RV, "

T. Hilage, R. Kulkarni, Application of data mining techniques to a selected business organization with special reference to buying behaviour", International Journal of Database Management Systems, Vol.3, No 4, pp.169-181, 2011

Y. Huang, S. Lin, Applying Multidimensional Association Rule Mining to Feedback-Based Recommendation Systems, IEEE International Conference on Advances in Social Network Analysis and Mining, ASONAM, July 2011, DOI: 10.1109/ASONAM.2011.29

S. Vinchukar, S. Nirkhi: Feature Extraction of Product from Customer Feedback through Blog, International Journal of Emerging Technology and Advanced Engineering, January 2012, (ISSN 2250-2459, Volume 2, Issue 1

U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, Knowledge Discovery and Data Mining: Towards a Unifying Framework, KDD-96 Proceedings, 1996

Open Refine, http://www.geobib.fr/blog/2020-04-29-openrefine-detect-lang, Access time: 25.8.2021

CS & S, Code for Science & Society, https://codeforscience.org/, Access time: 25.8.2021

Paperswitchcode, 2021, https://paperswithcode.com/task/multi-label-classification, Access time: 25.8.2021

H. Liu, C. Yuan, X. Wang, Label-Wise Document Pre-Training for Multi-Label Text Classification, NLPCC 2020, August 2020, https://arxiv.org/abs/2008.06695

M. Ibrohim, I. Budi, Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter, Proceedings of the Third Workshop on Abusive Language Online, August 2019, Florence, Italy, DOI: 10.18653/v1/W19-3506