# Analyzing Patient Feedback Data with Topic Modeling

*Citation for published version (APA):*
Arendsen, J., Rijcken, E., Zervanou, K., Rietjens, K., Vlems, F., & Kaymak, U. (2022). Analyzing Patient Feedback Data with Topic Modeling. In D. Ciucci, I. Couso, J. Medina, D. Ślęzak, D. Petturiti, B. Bouchon-Meunier, & R. R. Yager (Eds.), *Information Processing and Management of Uncertainty in Knowledge-Based Systems - 19th International Conference, IPMU 2022, Proceedings: 19th International Conference, IPMU 2022, Milan, Italy, July 11–15, 2022, Proceedings, Part II* (pp. 248-258). (Communications in Computer and Information Science; Vol. 1602 CCIS). Springer. https://doi.org/10.1007/978-3-031-08974-9_20

*DOI:*
10.1007/978-3-031-08974-9_20

*Document status and date:*
Published: 26/06/2022

*Document Version:*
Accepted manuscript including changes made at the peer-review stage

# Analyzing Patient Feedback Data with Topic Modeling

Jasper Arendsen[1], Emil Rijcken[1,2], Kalliopi Zervanou[3], Kim Rietjens[4], Femke Vlems[5], and Uzay Kaymak[1,2]

[1] Jheronimus Academy of Data Science, Den Bosch, The Netherlands
[2] Eindhoven University of Technology, Eindhoven, The Netherlands
[3] Leiden University, Leiden, The Netherlands
[4] Q-Qonsult Zorg, Utrecht, The Netherlands
[5] Antoni van Leeuwenhoek, Amsterdam, The Netherlands

**Abstract.** Patient feedback is an increasingly important measure to support quality improvement within healthcare organisations. Until recently, the focus has been on developing mechanisms for collecting patient feedback. However, research into analysis techniques to examine such feedback, especially free-text comments, is limited. The analysis of free-text data requires substantial effort because of the unstructured nature of the responses. As a result, this type of data is often underutilised within healthcare organisations while it contains the most valuable information. This research aims to analyse unstructured patient feedback, collected via a PREM questionnaire, utilising text mining. In particular, the extent to which topics can be extracted from this data is explored. Multiple topic modelling algorithms (LDA, FLSA, FLSA-W, NMF, BTM) are selected based on previous research and the data set characteristics. The applied topic modelling techniques proved to be able to provide a high-level overview of patient experiences. Hence, this research can be considered as one of the first steps towards automated analysis of unstructured patient feedback.

**Keywords:** Topic Modeling · Fuzzy Topic Models · Patient Feedback · Text Mining · Information Extraction.

## 1 Introduction

Over the years, patient feedback has become an increasingly important outcome measure for healthcare organizations, and it is one of the central pillars that supports quality improvement [8], [18]. The methods used to collect this data can be both quantitative and qualitative, and the obtained data can range from individual nurse-patient dialogues to standardized questionnaires [19]. The most common and structured method to collect patient experience data is via a Patient Reported Experience Measure (PREM). The PREM is a nationally coordinated method to measure the patient experience in hospitals [15]. Typically, a PREM contains both quantitative ratings and free-text fields focused on the

care provided in the hospital as experienced by the patient. The free-text fields allow the patient to elaborate on the ratings they provided. Also, they can provide information on experiences not covered by the questionnaire. Having such a better understanding of the patient allows hospitals to optimize the care they provide to the wishes of the patients and shift towards a more patient-centred healthcare service [23]. Moreover, understanding patients' specific dissatisfaction can help health professionals and administrators identify and rectify organizational deficiencies before they become costly [21]. The information captured in unstructured text fields may be very valuable for care improvement. However, research into the analysis techniques of free-field feedback is limited and often underutilized within the medical domain [5]; possibly because the analysis of unstructured data is challenging [9]. This research aims to analyze free-text patient feedback utilizing topic modeling. Firstly, we perform a grid search to optimize various topic modelling algorithms. Then, we evaluate the produced topics quantitatively and qualitatively (through domain experts). The domain experts found an in-depth analysis of the topics challenging due to its broad and ambiguous interpretation. Yet, the topics produced by the topic models do provide high-level insights.
The outline of the paper is as follows. In Section 2 we discuss the various topic modeling algorithms used in this research. In Section 3 we discuss our comparison methodology and data gathering and preprocessing. We present the results from both the quantitative and qualitative evaluation in Section 4. Then we discuss our findings in Section 5 and conclude the paper in Section 6.

## 2 Topic Modeling

A commonly used text mining method to analyze textual data is topic modeling. Topic modeling extracts hidden topics from a collection of documents. Although various algorithms exist, their output consists of two matrices:

1. $P(W_i|T_k)$- The probability of word $i$ given topic $k$,
2. $P(T_k|D_j)$- The probability of topic $k$ given document $j$

with:
$i$ word index $i \in \{1, 2, 3, ..., M\}$,
$j$ document index $j \in \{1, 2, 3, ..., N\}$,
$k$ topic index $k \in \{1, 2, 3, ..., C\}$,
$M$ the number of unique words in the data set,
$N$ the number of documents in the data set,
$C$ the number of topics.

Then, the $n$ words with the highest probability per topic are used to represent that topic.
The most popular algorithm is Latent Dirichlet Allocation (LDA), a generative probabilistic model. Documents are represented as random mixtures over latent

topics, where a distribution over words characterizes each topic [4]. The document length highly influences LDA's performance; it does not perform well on short texts [11], [26].

In contrast, Non-negative Matrix Factorization (NMF) is known to perform well on short texts [20]. This method projects high-dimensional vectors into a lower-dimensional space. It takes the document-word matrix and represents this in two matrices **U** and **V**. **U** consists of the topics found in the documents, and **V** consists of the corresponding coefficients representing the weights for those topics. **U** and **V** are calculated by optimizing the NMF objective function. NMF has fewer parameters than LDA and often distinguishes more realistic topics.

Another algorithm that is designed to deal with short text's sparsity is Biterm Topic Modeling (BTM) [25]. A biterm is defined as an unordered word-pair co-occurring in a short text. This method is based on the assumption that words occurring frequently together belong to the same topic.

Recently, Fuzzy Latent Semantic Analysis (FLSA) [13] was applied to healthcare data and showed superior results in comparison to LDA [4]. Just like NMF, FLSA starts with the document-term matrix. It then applies a global term weighting mechanism, after which the representation is projected onto a lower-dimensional space through singular value decomposition. Then, fuzzy C-means clustering [1] is performed on $\mathbf{U}^T$ (thus, documents are being clustered) to find different topics. Inspired by FLSA, FLSA-W works similarly but clusters on $\mathbf{V}^T$ and thus, clusters on words [22]. It outperforms both LDA and FLSA in experimental studies.

## 3 Analyzing Patient Feedback

We aim to analyse unstructured patient feedback using topic modeling methods. In particular, we explore the extent to which topics can be extracted from this data consisting of Dutch texts. The main steps to achieve this goal are shown in Figure 1. Below, we explain each step.

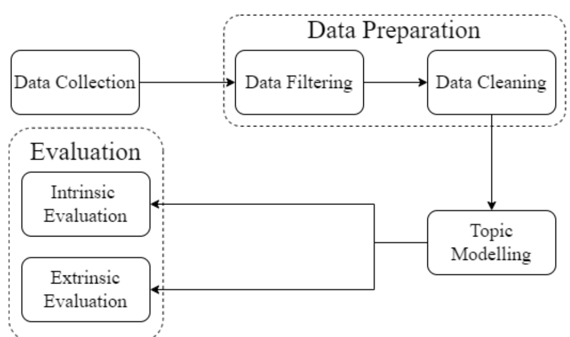


**Fig. 1.** Overview of the study methodology

### 3.1 Data Collection

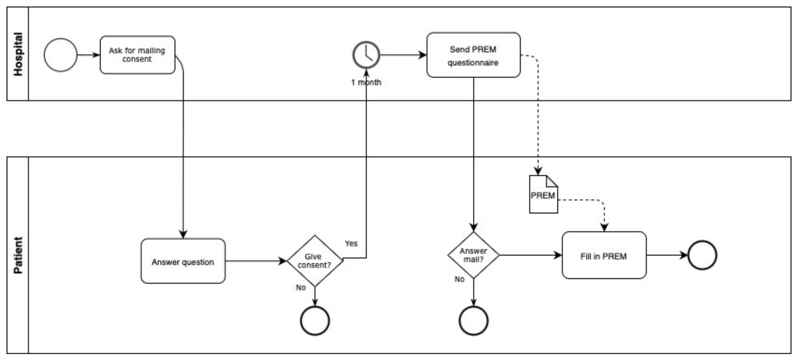


**Fig. 2.** PREM Data Collection Process.

The data consists of the feedback given in the PREM, collected by the Antoni van Leeuwenhoek (AVL) hospital, a hospital and research institute located in Amsterdam in The Netherlands. There are two types of patients (patients with a malign tumour and patients with a benign tumour) who are asked to fill out different questionnaires because of the difference in treatment. Since most patients have a malign tumour and this type of care is the most important for the AVL, we consider only their feedback in this study. The PREM questionnaire consists of three categories of questions. Firstly, patients are asked for practical data about their treatment. This includes questions about their type of illness, contact, treatment type and whether they participated in a trial. Secondly, patients are asked to rate their satisfaction regarding the provided care overall and for individual healthcare components. The individual components are Interaction, Relationship with healthcare providers, Expertise, Atmosphere, Waiting times, Available information, Aftercare, Facilities, Research, Parking, and others. Lastly, there are two open-ended questions in which the patient can expand on ratings given earlier to indicate what they were and were not satisfied with. In this study, we analyze these two fields with unstructured data separately and refer to them as 'satisfaction data' and 'improvement data'.

### 3.2 Data Preprocessing

The data preparation phase involves two main steps: (i) data filtering and (ii) data cleaning. Both steps are described in this subsection. Three ways of filtering are applied to the patient feedback data:

1. non-malign patients are removed,
2. missing values are removed.

3. non-Dutch entries are removed [6].

For data cleaning, we have performed the following steps to the filtered data:

1. string cleaning [7]
2. tokenization,
3. lowercasing,
4. stopword removal, [8]
5. punctuation removal,
6. hex-digit removal,
7. lemmatisation [9].

### 3.3 Topic Modeling

Several parameters need to be set to train and optimize topic models. Two parameters are set for all algorithms: the number of topics and the number of words per topic. Additionally, each algorithm has its own set of hyperparameters to be tuned. The optimal number of topics are five and six, for the satisfaction and improvement data, respectively. They are found by using the elbow method [14] [16] for determining an acceptable trade off between maximizing topic coherence (see below) and minimizing the number of topics. This number stays fixed while tuning the other hyperparameters. Furthermore, only the top 10 words per topic are selected in order to reduce the required time effort for the qualitative evaluation by the domain experts. Table 1 shows the optimal- and grid search values for the model-dependent hyperparameters. In this table, the symmetric and asymmetric values in lda's range refer to the Dirichlet priors used [24].

### 3.4 Evaluation

Evaluation methods of topic model quality can be divided into (i) intrinsic and (ii) extrinsic methods. Intrinsic evaluation methods rely on internal evaluation metrics which directly quantify the task performance, while extrinsic evaluations focus on external evaluation metrics.

**Intrinsic Evaluation** Since a topic model's output consists of various topics, each topic containing a collection of words, the quality of a topic model should focus both on the quality of words within each topic (intra-topic quality) and the diversity amongst different topics (inter-topic quality). For the intra-topic quality, we use the $C_v$ coherence metric, which correlates highest to human

[6] Filtering is done by applying the Python langdetect algorithm [6]. This algorithm supports over 50 languages and has a precision of 99.77%.

[7] Caused by a different data encoding in the hospital's database.

[8] Stop words are removed from the data by implementing the NLTK Dutch package. [3]

[9] Lemmatisation is applied using the Spacy Dutch Python package [12].

**Table 1.** Hyperparameter Grid Search Settings and Optimal Values

| Model | Parameter | Range | Optimal Value Satisfaction | Optimal Value Improvement |
|---|---|---|---|---|
| **lda** | $\alpha$ | [0.01, 0.2, 0.4, 0,6, 0.8, 1, *symmetric*, *asymmetric*] | 1 | 0.4 |
| | $\beta$ | [0.01, 0.2, 0.4, 0,6, 0.8, 1, *symmetric*, *auto*] | 0.4 | 0.8 |
| | eval_every | [5-40, *steps = 5*] | 5 | 25 |
| | passes | [1-15, *steps = 1*] | 14 | 9 |
| **flsa** | word_weighting | [idf, probidf] | probidf | probidf |
| | cluster_method | [fcm, fst-pso, gk] | fst-pso | fst-pso |
| | svd_factors | [1-5] | 4 | 4 |
| **flsa-w** | word_weighting | [idf, probidf] | probidf | probidf |
| | cluster_method | [fcm, fst-pso,gk] | fst-pso | fst-pso |
| | svd_factors | [1-5] | 4 | 4 |
| **nmf** | kappa | [0.01, 0.25, 0.5, 0.75, 1, 2] | 0.5 | 1 |
| | eval_every | [5-40, *steps = 5*] | 20 | 15 |
| | passes | [1-15, *steps = 1*] | 10 | 15 |
| **btm** | $\alpha$ | [0.01, 0.2, 0.4, 0,6, 0.8, 1] | 0.8 | 0.2 |
| | $\beta$ | [0.01, 0.2, 0.4, 0,6, 0.8, 1] | 1.0 | 0.2 |
| | iterations | [100-800, *steps = 100*] | 100 | 200 |

judgment amongst all coherence metrics. With $c_v$, the Normalized Pointwise Mutual Information (2) is calculated for the combination of all the top-$n$ words in a topic. Then, the arithmetic mean is calculated based on all these scores. To calculate the probabilities in Normalized Pointwise Mutual Information, a sliding window of 110 words is being used.

$$PMI(w_i, w_j) = log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)} \tag{1}$$

$$NPMI(w_i, w_j)^\gamma = \frac{PMI}{\sum_{i=1}^{M} \sum_{j=1}^{N} P(W_i, D_j)}^\gamma \tag{2}$$

The coherence score ranges between zero and one, where one means perfect coherence and zero means no coherence whatsoever.
Additionally, the $C_{UMass}$ is considered, which has the second-highest correlation with human judgment. We use the diversity score for the inter-topic quality. We define the following quantities:

- $W_{unique}$ the number of unique words in the top-$n$ words of all the topics,
- $W_{all}$ the total number of words in all the topics ($n \times C$),
- $n$ the number of words per topic,
- $C$ the number of topics.

Then, Equation 3 shows how diversity is calculated.

$$Diversity = \frac{W_{unique}}{W_{all}}. \tag{3}$$

Lastly, we calculate the interpretability score as the product between the coherence ($C_v$) and diversity score [7], as can be seen in equation 4.

$$Interpretability = Coherence(c_v) \times Diversity \tag{4}$$

**Extrinsic Evaluation** The domain experts are presented with three questions per topic to measure both the inter- and intra-topic quality for the extrinsic evaluation. For the intra-topic quality, they are asked to rate the topics on their individual quality, similar to [2]. The quality is communicated to the experts as a combination of the coherence, meaningfulness, and interpretability of the top $n$ words per topic with respect to their weights. The quality of the individual topics is measured via an ordinal score [0-3], where 3 represents a 'good/useful' topic, and 0 defines a 'bad/useless' topic. Furthermore, the domain experts are asked to assign each topic with one of the categories extracted from the PREM for the inter-topic quality. For extrinsic evaluations, it is important to take the inter-rater agreement into account because of the variation in human interpretation. Since this research includes nine raters, the Krippendorff's alpha score $\alpha$ is applied to indicate the interrater agreement [17]. Four raters focused on the satisfaction data and five on the improvement data to save time and costs.

## 4 Results

### 4.1 Intrinsic Evaluation

The intrinsic metrics, discussed in Section 3.4, are presented in Table 2 and 3, where the best values per metric are boldface. The improvement data yields higher coherence scores than the satisfaction data, while diversity scores are comparable for both data sets. FLSA-W and FLSA perform best for both the satisfaction and improvement data. FLSA has the highest coherence scores, whereas FLSA-W produces the most diverse topics. As a result, FLSA-W produces the most interpretable topics for the satisfaction data and FLSA the most interpretable topics for the improvement data. Furthermore, LDA performs much worse than the fuzzy algorithms.

### 4.2 Extrinsic Evaluation

The Krippendorff's alpha score for the satisfaction data is 0.046, and the alpha score for the improvement evaluation is 0.042. These scores indicate that the results can be interpreted as statistically unrelated. The extrinsic evaluation

**Table 2.** Model Intrinsic Values of the Satisfaction Data. With $C_v$ and $C_{U_{Mass}}$ coherence score

| | Satisfaction | | | |
|---|---|---|---|---|
| 2-5 Model | $C_v$ | $C_{UMass}$ | Diversity | Interpretability |
| **lda** | 0.490 | -2.379 | 0.780 | 0.382 |
| **flsa** | **0.688** | **-2.335** | 0.880 | **0.605** |
| **flsa-w** | 0.547 | -5.384 | **0.920** | 0.503 |
| **nmf** | 0.591 | -2.590 | 0.860 | 0.508 |
| **btm** | 0.518 | -3.216 | 0.820 | 0.425 |

**Table 3.** Model Intrinsic Values of the Improvement Data. With $C_v$ and $C_{U_{Mass}}$ coherence score

| | Improvement | | | |
|---|---|---|---|---|
| 2-5 Model | $C_v$ | $C_{UMass}$ | Diversity | Interpretability |
| **lda** | 0.575 | -2.134 | 0.567 | 0.326 |
| **flsa** | **0.770** | **-1.449** | 0.883 | 0.680 |
| **flsa-w** | 0.755 | -2.269 | **0.950** | **0.717** |
| **nmf** | 0.760 | -2.074 | 0.900 | 0.684 |
| **btm** | 0.632 | -2.263 | 0.733 | 0.463 |

scores are shown in Table 4. The 'mean' score shows the average ordinal score [0-3] as assigned by the experts. The 'Uniqueness' indicates the average number of labels a topic was assigned to. For this value, a higher value indicates its more challenging to assign a label to a topic, and the ideal value is one. This could mean a topic is hard to interpret. The 'No Category' is used when the human subjects cannot fit the topic into any category. Ideally, this value is as low as possible. Since the variation in scores is relatively small, it seems that quality differences between the topic models are relatively low. Nonetheless, the algorithms generally score higher on the satisfaction data than the improvement data. LDA has the best quality topics for the satisfaction data, whereas FLSA performs best on the improvement data. The differences in performance between the datasets is likely caused by the dataset's characteristics; the texts from the improvement data are longer and more unique, generally, than the satisfaction data.

## 5 Discussion

After evaluating the topics both intrinsically and extrinsically, we find contradicting results between the two. According to the intrinsic evaluation, the improvement dataset has the highest intra-topic quality, whereas the satisfaction scores best according to the extrinsic evaluation. The results are still rather preliminary and We need to conduct further experiments to better understand which evaluation has the most impact. The difference between the two metrics is

**Table 4.** Model Extrinsic Values

| | Satisfaction | | | Improvement | | |
|---|---|---|---|---|---|---|
| Model | Mean | Uniqueness | No Category | Mean | Uniqueness | No Category |
| **lda** | **1.550** | 2.300 | 0.000 | 1.167 | 1.467 | 0.233 |
| **flsa** | 1.300 | 2.300 | 0.050 | **1.300** | 1.533 | 0.233 |
| **flsa-w** | 1.500 | 2.350 | 0.000 | 0.967 | 1.300 | 0.300 |
| **nmf** | **1.550** | **2.450** | 0.000 | 1.267 | 1.600 | 0.133 |
| **btm** | 1.250 | 1.900 | 0.000 | 1.233 | **1.700** | 0.200 |

likely due to the different characteristics of both datasets: the satisfaction data has a lower word variability than the improvement data. However, this may also be caused by the low inter-rater reliability in the extrinsic evaluation. Generally, the domain experts find the quality of the topics relatively low due to their broad and ambiguous interpretation. In particular, the mean quality is perceived as relatively low, and the uniqueness and fraction of no categories were relatively high. Although differences within the extrinsic evaluation were low, dissimilarity is noticeable concerning the satisfaction and improvement results. Satisfaction topics are interpreted more broadly due to the average number of categories selected. Improvement topics are more ambiguous as the fraction of no fitting topics is relatively high. As a result of the broad and ambiguous interpretation, the models only allow for a high-level topic overview.

In this research, we have not considered the quantitative scores given by patients, while these scores are likely to provide valuable insights. Additionally, the Krippendorff's alpha scores, used for the extrinsic evaluation, are close to zero, indicating an absence of inter-rater agreement. The low agreement between the raters affects the reliability of the extrinsic evaluation. Even so, the low alpha score can be caused by the possible subjective interpretation of the results or the low number of domain experts [10].

Finally, we use the elbow method to determine the optimal number of topics. A typical 'elbow' is noticeable for the satisfaction data. Yet, no such pattern can be found with the improvement data. Hence, a suboptimal number of topics might have been selected. Consequently, this method is not ideal for determining the number of clusters for the improvement corpus. Therefore, a more detailed data analysis should determine the optimal number of topics.

## 6 Conclusion

In this work, we analyze free-text patient feedback to find relevant information to improve healthcare practices. We have trained/optimized various topic modeling algorithms and evaluated both intrinsically and extrinsically. FLSA and FLSA-W have the highest intrinsic scores, whereas NMF and BTM perform best on the extrinsic evaluation. The methodology used in this paper can be implemented into the hospital's dashboard so that patient feedback is monitored more regularly and adequately. In future work, we plan to include more cohorts of patients to assess the generalizability of our results.

## References


1. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Springer Science & Business Media (2013)
2. Bhatia, S., Lau, J.H., Baldwin, T.: An automatic approach for document-level topic model evaluation. In: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017). pp. 206–215. Association for Computational Linguistics, Vancouver, Canada (Aug 2017). https://doi.org/10.18653/v1/K17-1022, https://aclanthology.org/K17-1022
3. Bird, S., Klein, E., Loper, E.: Natural Language Processing With Python: Analyzing Text With the Natural Language Toolkit. " O'Reilly Media, Inc." (2009)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. The Journal of Machine Learning Research **3**, 993–1022 (2003)
5. Cunningham, M., Wells, M.: Qualitative Analysis of 6961 Free-Text Comments From the First National Cancer Patient Experience Survey in Scotland. BMJ open **7**(6), e015726 (2017)
6. Danilk, M.: Langdetect. PyPI (2014), https://pypi.org/project/langdetect/
7. Dieng, A.B., Ruiz, F.J., Blei, D.M.: Topic modeling in embedding spaces. Transactions of the Association for Computational Linguistics **8**, 439–453 (2020)
8. Doyle, C., Lennox, L., Bell, D.: A Systematic Review of Evidence on the Links Between Patient Experience and Clinical Safety and Effectiveness. BMJ Open **3**(1) (2013). https://doi.org/10.1136/bmjopen-2012-001570, https://bmjopen.bmj.com/content/3/1/e001570
9. Gleeson, H., Calderon, A., Swami, V., Deighton, J., Wolpert, M., Edbrooke-Childs, J.: Systematic Review of Approaches to Using Patient Experience Data for Quality Improvement in Healthcare Settings. BMJ Open **6**(8) (2016). https://doi.org/10.1136/bmjopen-2016-011907, https://bmjopen.bmj.com/content/6/8/e011907
10. Hallgren, K.A.: Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutorials in Quantitative Methods for Psychology **8**(1), 23 (2012)
11. Hong, L., Davison, B.D.: Empirical Study of Topic Modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics. pp. 80–88 (2010)
12. Honnibal, M., Montani, I.: spaCy 2: Natural Language Understanding With BLoom Embeddings, Convolutional Neural Networks and Incremental Parsing (2017), to appear
13. Karami, A., Gangopadhyay, A., Zhou, B., Kharrazi, H.: Fuzzy Approach Topic Discovery in Health and Medical Corpora. International Journal of Fuzzy Systems **20**(4), 1334–1345 (2018)
14. Khalid, H., Wade, V.: Topic Detection from Conversational Dialogue Corpus with Parallel Dirichlet Allocation Model and Elbow Method. In: Computer Science & Information Technology. AIRCC (2020)
15. Kingsley, C., Patel, S.: Patient-reported outcome measures and patient-reported experience measures. Bja Education **17**(4), 137–144 (2017)
16. Kodinariya, T.M., Makwana, P.R.: Review on Determining Number of Cluster in K-Means Clustering. International Journal **1**(6), 90–95 (2013)

17. Krippendorff, K.: Content Analysis: An Introduction to Its Methodology. Sage Publications (2018)
18. Kumah, E., Ankomah, S.E., Kesse, F.O.: The Impact of Patient Feedback on Clinical Practice. British Journal of Hospital Medicine **79**(12), 700–703 (2018)
19. LaVela, S.L., Gallan, A.: Evaluation and Measurement of Patient Experience. Patient Experience Journal **1**(1), 28–36 (2014)
20. Lee, D.D., Seung, H.S.: Learning the Parts of Objects by Non-negative Matrix Factorization. Nature **401**(6755), 788–791 (1999)
21. Pichert, J.W., Miller, C.S., Hollo, A.H., Gauld-Jaeger, J., Federspiel, C.F., Hickson, G.B.: What Health Professionals Can Do to Identify and Resolve Patient Dissatisfaction. The Joint Commission Journal on Quality Improvement **24**(6), 303–312 (1998). https://doi.org/https://doi.org/10.1016/S1070-3241(16)30382-0, https://www.sciencedirect.com/science/article/pii/S1070324116303820
22. Rijcken, E., Scheepers, F., Mosteiro, P., Zervanou, K., Spruit, M., Kaymak, U.: A comparative study of fuzzy topic models and lda in terms of interpretability. In: Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence (SSCI). pp. 1–8 (2021)
23. Shaller, D.: Patient-Centered Care: What Does It Take? Commonwealth Fund New York (2007)
24. Syed, S., Spruit, M.: Exploring symmetrical and asymmetrical dirichlet priors for latent dirichlet allocation. International Journal of Semantic Computing **12**(03), 399–423 (2018)
25. Yan, X., Guo, J., Lan, Y., Cheng, X.: A Biterm Topic Model for Short Texts. In: Proceedings of the 22nd International Conference on World Wide Web. pp. 1445–1456 (2013)
26. Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., Li, X.: Comparing Twitter and Traditional Media Using Topic Models. In: European Conference on Information Retrieval. pp. 338–349. Springer (2011)