

What are they telling us? Accessible analysis of free text data from a national survey of higher education students

Sean O'Reilly¹ and Geraldine Gray²

¹ The Technological Higher Education Association, Fumbally Square, Dublin 8, Ireland

² Technological University Dublin, Blanchardstown Road North, Dublin 15, Ireland

Abstract

Many staff in higher education have a sense that useful information is buried within their data that they are unsure how to access, or even what questions it can answer. This is particularly so with survey text responses from large student cohorts. This paper examines valid and repeatable methods to analyze such data while seeking to minimize computational and analyst workload by maximizing machine learning to accommodate the large volume of data. We evaluate clustering and topic modelling as methods to analyze one year's data from a national student survey in Ireland, an anonymized dataset with more than 44,700 respondents. The primary focus was on free text responses to two questions, namely those seeking to identify the best aspects of students' reported experiences, and those identifying aspects that need improvement. K-means and Latent Dirichlet Allocation unsupervised learners were used to identify key themes emerging from the text data. K-means proved computationally expensive and failed to usefully categorize significant minorities of the data. In contrast, topic modelling had relatively low overheads and effectively categorized more than 97% of the sample data into themes which could be usefully considered in the business domain. From this research, topic modelling provided an effective method to analyze such text data once careful consideration was given to determining the appropriate initial number of topics for configuring the algorithm.

Keywords

Higher education, student survey, free text, machine learning, unsupervised, clustering, topic modelling, k-means, LDA

1. Introduction

Surveys of students' experiences have become very common in many higher education systems as the collection and utilization of data for a variety of practical and policy based purposes has been steadily increasing [1]. However, the predominant focus of analysis has tended to be on quantitative responses, largely due to the analytical challenges of qualitative data analysis. This paper reports on analysis of text responses collected from one iteration of fieldwork for the Irish Survey of Student Engagement [2], an annual national survey of higher education students in Ireland. The survey included two questions seeking free text responses which asked students to report on the best aspects of their experiences and on those aspects that needed improvement.

This research sought to address the apparent lack of both understanding and capacity to efficiently analyse qualitative data generated by the survey. This was done by exploring multiple analytical approaches to identify effective methods which could, in due course, be shared with stakeholders to encourage and promote such analysis more widely. Therefore, the research question asked *what is the most efficient method to analyse text data from open survey responses?* The main objective set at the outset was to identify a valid and repeatable method which sought to minimize analyst and

Proceedings of the Finnish Learning Analytics and Artificial Intelligence in Education Conference (FLAIEC22), Sep 29–30, 2022, Joensuu, Finland

EMAIL: Sean.OReilly@thea.ie (A. 1); Geraldine.Gray@tudublin.ie (A. 2)

ORCID: 0000-0002-9694-2558 (A. 2)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

computational workloads to inform dissemination and promotion to survey partners. A selection of methods identified during background research were assessed.

This paper provides an overview of the dataset, which was relatively large in the context of Irish higher education, and reports on the research carried out to respond to the research question and objective through an iterative process of experimentation and model implementation. Section 2 provides an overview of background research undertaken. Section 3 describes the dataset involved and explains the methods selected. Section 4 presents headline results achieved from a series of experimental models. Section 5 offers analysis of those results and notes a number of possible areas for future research. Finally, Section 6 presents the conclusions.

2. A review of methods for analysing open survey responses

Despite the “bewildering variety of strategies” for text mining available to the analyst [3], background research identified considerable consensus on the basic steps required to prepare text data for analysis. Almost all papers reported: the use of tokenisation; the removal of punctuation and other non-letter characters; conversion to lower case; and creation of a document matrix, for example [4-7]. Document vectors using term frequency - inverse document frequency (TF-IDF) as the occurrences count were regarded as most effective as they provided an insight into the relative importance of tokens to the overall body of text [8]. Differing views were reported on the impact of stemming and the removal or retention of stop words [4,6,9]. The impact of using n-grams was discussed less frequently and typically in the context of sentiment analysis, e.g [6]. Part of speech (POS) tagging was reported relatively rarely but with the key purpose of reducing the dimensionality of the data set prior to further analysis.

Clustering techniques featured as feasible methods to analyse survey text data in a number of publications. For example, [6] used agglomerative clustering in their Student Feedback Mining System (SFMS) to analysis survey responses. The importance of visualisations of hierarchical cluster models and the limitation that each document could belong only to a single cluster was noted by [5]. Topic modelling was identified as a method which could address that limitation, with Latent Dirichlet Allocation (LDA) particularly wellsuited to short texts and a working assumption that documents have multiple topics [10,11]. The advantage of not having to label a set of documents in advance because of the use of an unsupervised classifier such as LDA was argued by [11]. They also concurred with [10] that determining an optimal number of topics in advance of detailed analysis was a potential limitation. [12] offered valuable insights into configuration of LDA parameters and, importantly, observed that the number of documents was more important than their length. These researchers also reported that topic modelling could not replace qualitative coding of text but would provide additional information as a complementary tool. These findings proved informative in planning implementation and experimentation for this research which made use of a large number of, typically, short responses which ultimately were read by the analyst to evaluate the quality of topics. Issues of evaluating the performance of topic modelling were raised by [13] who found that the optimal number of topics as suggested by the intrinsic measure of perplexity was often inversely correlated with human judgement; and, further, that in this scenario, human judgement should win the argument.

While classification has been successfully used to label students’ open survey responses, such as in [14], it necessitates that those topics are defined in advance, and so limits the scope of discoverable topics. This, along with the limitation of one topic per response, and the lack of labelled data available for this project, meant classification on its own was not an option. However, [15] successfully used Naïve Bayes to evaluate the predictability of cluster membership.

3. Methodology

The dataset used originated from the 2020 iteration of fieldwork for the national survey of students’ engagement in higher education. Responses were collected in February and March 2020, and so predominantly reflect perspectives prior to COVID restrictions [2]. The dataset had been anonymized by the external survey contractor prior to return to survey partners and was further anonymized

prior to receipt for this project, to remove identifiers for individual institutions. The dataset consisted of responses from 44,707 students across 142 attributes.

This represented 31% of the target student population across a broad range of disciplines from a variety of higher education institutions, including all traditional universities, technological universities, institutes of technology, colleges of education and a number of private institutions.

Attributes could be grouped into 3 broad categories: demographic data (such as year group, ISCED field of study, mode of study, etc); actual question responses; and attributes calculated from the prior two groups. The results from analysis of quantitative responses at national level varied little from year to year since the survey was first administered in 2014, with greatest variation in results within individual institutions [2]. However, there has been very little analysis of survey free text data in the public domain. That fact alone provided the context for undertaking this study. Analysis of the national dataset was necessarily heavily reliant on automated and replicable processes because of the volume of data. This focus meant that a certain degree of inaccuracy was regarded as acceptable because of the relatively “high” level at which analysis was undertaken, as for [16] and [17]. It is acknowledged that analysis of survey text data within individual institutions would be of greatest value when automated analysis was complemented by more detailed qualitative analysis [12].

Informed by the background research, efforts to address the research question were necessarily explorative in nature, taking a number of feasible approaches adopted elsewhere, to explore and analyse the data and evaluate the effectiveness of the approaches taken up to that stage, given the overall business context. The published papers reviewed tended to each focus on a small number of analysis methods and to report on their findings, rather than compare different approaches. It appeared that little research had been undertaken to compare different methods in order to determine which may offer efficient ways to analyse text data and, therefore, that this study may provide new information to others working with text data, particularly text data generated from student surveys.

3.1. The study dataset

The priority focus for this research was on responses to two specific survey questions. These asked “What does your institution do best to engage students in learning?” and “What could your institution do to improve students’ engagement in learning?”. Throughout this paper hereafter, these are referred to as Best Aspects (BA) and Needs Improvement (NI). In the original dataset, 18,494 rows contained values for BA data and 20,205 rows contained values for NI data. An overview of these data after removal of non-letters, single letters and blanks is provided in Table 1. As indicated by the differences between mean and median lengths, a large proportion of responses were short – with, for example, 10,216 BA responses and 7,784NI responses each containing less than 50 characters.

3.2. Data Preparation

The data was converted to individual text documents with filenames which included a key identifier and a label indicating best aspects, BA, or needs improvement, NI. Based on consensus from background research, the data was transformed to lower case and tokenized (using non-letters). Shorter tokens which contained meaning in the business context were replaced by full titles to ensure that such information was retained when shorter tokens were subsequently filtered. Examples of these replacements included SU (student union), CA (continuous assessment), and MCQ (multiple choice quiz). A bespoke dictionary was used to list terms to be filtered out from the text data. These included acronyms and other terms which clearly identified individual institutions. Tokens were also filtered by length to remove those with less than 4 characters. Document vectors used TF-IDF to count term frequencies. Data sets were pruned to remove infrequent words (in less than 3% of comments) and frequent words (in more than 30% of comments). These values were chosen as a result of exploring the effect of different values to prune the dataset sufficiently to bring computational reductions without notable loss of potential information value. Experimenting with bigrams resulted in computation times of multiples of hours (>44) but failed to create meaningful clusters. A variety of levels of term pruning and numbers of clusters resulted in >85% of comments allocated to one large cluster in all cases. Therefore, bigrams were deemed unfeasible.

Unigram datasets were prepared with and without Porter stemming, and with retention and removal of stop words to accommodate exploration of differing findings from background research. Data was also divided into 70% training and 30% testing to support evaluation of models developed. This was done by repeatedly taking arbitrary groups from the file in order to reduce the risk of any unknown sequencing impacting future results. In hindsight, random sampling may have been a better approach to increase the level of automation for deployment.

Table 1.
Summary of descriptive statistics for free text data

Attribute	Min length	Max length	Mean length	Median length	St dev
(BA)	2	1273	49.2	27	61.4
(NI)	2	3281	73.5	38	117.6

3.3. Chosen modelling approaches

The research question asked about the most efficient method to analyze the text data responses. Based on background research, clustering and topic modelling were examined in some detail with the objective of comparing their effectiveness, as had been undertaken to some extent with scientific documents by [11]. The project objectives of focusing on machine learning and seeking to identify a “one-off” method, rather than developing and subsequently applying a coding frame or manually labelling training data, meant that unsupervised learners offered the preferable solutions. The importance of determining the initial number of clusters and topics had been highlighted in multiple published papers, so this issue was explored in detail.

3.3.1. Cluster modelling

Background research identified the potential of clustering to identify key themes, both hierarchical [5] and agglomerative [6]. It was anticipated that this approach may be insufficient alone to address the potentially multiple issues identified in responses to the two prompt questions, whereas topic modelling using Latent Dirichlet Allocation (LDA) assumed that documents had multiple topics, as also noted by [6]. Accordingly, clustering was explored with a finite number of options to determine its ability to meet the business need of identifying frequently occurring themes. This approach was deemed suitable to the business context for this research which sought to minimise analyst ‘manual’ input to identify an efficient ‘one-off’ approach.

Initial experiments with agglomerative, x-means and k-means cluster models using the full dataset identified that computational cost was a major factor to consider in order to meet the business objective of achieving an “efficient” analysis process. Therefore, a “fast” version of k-means that used triangle inequality to accelerate (standard) k-means was chosen for more detailed evaluation of clustering as a method [18]. Following the methodology used in [15], K-means models were run for a reduced (10%) dataset of BA and NI responses initially to assess number of topics. All values of k from 9 to 40 were assessed using Davies-Bouldin index. This index is a ratio between cluster scatter (within-cluster distances) and separation between clusters (between-cluster distances) and so, lower values were regarded as indicating better clusters. Modelling BA and NI together, optimal values for k were found at 11 and 34, with a marginal higher local optimum at k=26. When modelling BA and NI texts separately, each had a single optimal value at k=27 for BA and k=29 for NI. The results section will describe the results of these k-mean configurations trained on the full training dataset. All models were initially evaluated by using Naïve Bayes to predict the cluster(s) allocated to the training dataset by the methods described above, using a simple holdout of 40% for validation. The best performing models, i.e. those with greatest accuracy of Naïve Bayes label predictions and proposed clusters were further evaluated by applying k-means and Naïve Bayes models to the 30% unseen (test) dataset, prepared identically with the same list of terms.

3.3.2. Topic Modelling

The frequently referenced Latent Dirichlet Allocation (LDA) algorithm was used for topic modelling. As was the case for clustering, a key question to be considered was the initial parameter setting for number of topics. Similarly to k-means above, this was explored by executing multiple iterations of the LDA operator over the same range of number of topics. Perplexity was reported in multiple papers as a commonly used measure for evaluation of topics. It is calculated as the inverse of the mean per-word likelihood. However, background research had identified that perplexity and human judgement were not necessarily aligned so coherence was also considered [19]. Topic coherence measured the extent to which high scoring words in the same topic were related semantically and so offered one option to differentiate between topics which were “only” statistically sound and those which were likely to be more interpretable semantically. Multiple iterations of logging values of perplexity and coherence for ranges of numbers of topics consistently found lowest perplexity at the maximum number of topics examined. Therefore, “good” local maxima for coherence were used to develop a series of models on the full training dataset that were subsequently evaluated using Naïve Bayes to predict the proposed topic. It was noted that topic modelling produced a series of confidence values which enabled multiple topics to be related to single documents and, therefore, Naïve Bayes’ accuracy as measured by a single correct prediction did not reflect an entirely accurate view of the usefulness of proposed topics.

3.4. Implementation details

There are several possible approaches to minimize computational overheads and many of these require specific expertise for efficient use. This research used RapidMiner Studio (version 9.9), an open-source application with a readily understood graphical user interface which made it suitable for this research. A number of datasets were prepared for modelling from the same core data, as discussed in Section 3.2. These TF-IDF vectors were created using multiple configurations of the Process Documents from Files operator i.e., with various combinations of RapidMiner operators to: Transform Cases to lower case; Tokenise using non-letters; Filter Stopwords (or not); and apply Porter Stemming (or not), as shown in Figure 1. For topic modelling as discussed in Section 3.3.1, the Optimize Parameter operator was used to explore potentially optimal numbers of topics for further use in analytical models, as shown in Figure 2. This enabled testing different values for ‘number of topics’ in the embedded Extract Topics from Data (LDA) operator. Resulting perplexity and coherence values informed the selection of a limited number of topics for further analysis. This was done as a series of single models applied to specific prepared datasets as shown in Figure 3. In each case, the saved models which appeared to offer better performance were applied to the test dataset.

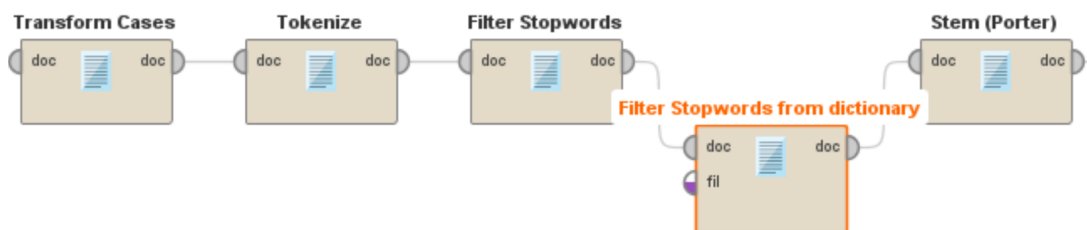


Figure 1: The data preparation processes nested within the *Process Documents* operator.

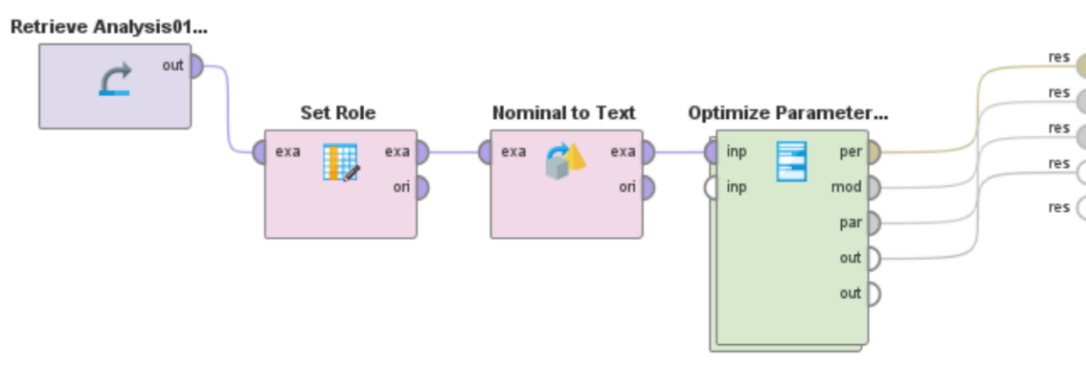


Figure 2: Using *Optimise Parameter* to identify potentially informative values for *numbers of topics*

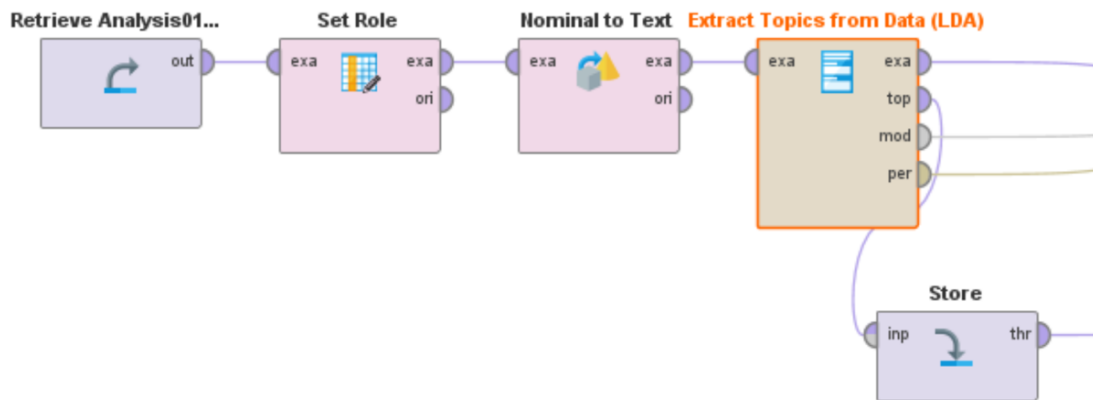


Figure 3: Storing results of the *LDA* operator with potentially useful *number of topics*

4. Results

4.1. Clustering with k-means

Acknowledging the risk of identifying only local minima for Davies-Bouldin, a series of clustering models were developed using the suggested “good” values for the number of clusters from 10% of the data, namely $k=11, 26, 27, 29$ and 34 . These were applied to the full training dataset(s) with / without stemming and with / without stop words. The models were initially evaluated by setting the proposed cluster names as labels and using a Naïve Bayes classifier to predict class labels. Results presented in Table 2 represent the optimal results based on Naïve Bayes accuracy for both BA and NI. The most accurate clustering models were achieved without stemming and with removal of stop words for 27 clusters of BA and for 29 clusters of NI data. The accuracy as reported for Naïve Bayes of applying these models to unseen (test) data was remarkably high, 99.22% for BA and 85.66% for NI data.

Further evaluation of the content of these clusters is discussed in Section 4.3. The distribution of examples to clusters is discussed in Section 5.

4.2. Topic modelling with Latent Dirichlet Allocation

Table 3 demonstrates that the most accurate LDA models were found with stemming, matching background research. The best performing models had relatively low numbers of topics, 8 topics for BA data and 10 topics for NI data. This may reflect that topic modelling allowed for multiple topics whereas clustering sought to identify themes for individual clusters and, so, the best performing clustering models involved notably higher numbers of clusters. The three best “performing” models for BA and for NI datasets from Table 3 were then applied to unseen (test) data for BA and for NI, which resulted in a notable drop in accuracy as illustrated in Table 4. As noted in Section 3.3.2, a limitation of classification model accuracy is that it is based on predicting one topic per statement.

Table 2

Results of Naive Bayes prediction of k-means clusters for BA and NI data

Dataset	Stemming?	Filter Stop words?	Value of k	Naïve Bayes Accuracy %	Clusters sizes* (decimal percentage)
BA	No	No	27	93.76	1658 (0.394) to 14 (0.003)
		Yes		81.10	1156 (0.275) to 2 (0.000)
	Yes	No	90.53	1243 (0.295) to 41 (0.010)	
		Yes	75.58	966 (0.229) to 7 (0.002)	
NI	No	No	29	92.16	1380 (0.372) to 23 (0.006)
		Yes		58.39	781 (0.210) to 63 (0.017)
	Yes	No	85.81	1099 (0.296) to 1 (0.000)	
		Yes	50.92	612 (0.165) to 5 (0.001)	

*Actual numbers for each cluster reflect a 60:40 data split to train Naïve Bayes and estimate model accuracy. The decimal percentage is most telling for relative distribution.

Table 3.

Results of Naïve Bayes prediction of LDA topics (clusters)

Training subset	Number of topics	Stemming?	Naïve Bayes Accuracy (%)	Clusters sizes* (decimal percentage)
BA	7	No stem	51.97	1491 (0.354) to 112 (0.027)
		Stem	57.59	1718 (0.408) to 157 (0.037)
	9	No stem	55.33	1931 (0.459) to 53 (0.013)
		Stem	56.71	1100 (0.261) to 107 (0.025)
	15	No stem	44.68	1210 (0.287) to 9 (0.002)
		Stem	49.64	1458 (0.346) to 35 (0.008)
	39	No stem	19.61	1605 (0.381) to 0 (0.000)
		Stem	18.25	1805 (0.429) to 0 (0.000)
NI	8	No stem	45.82	1207 (0.325) to 100 (0.027)
		Stem	40.13	1514 (0.408) to 57 (0.015)
	10	No stem	38.34	1707 (0.460) to 9 (0.002)
		Stem	44.5	1042 (0.281) to 88 (0.024)
	14	No stem	30.85	1653 (0.445) to 16 (0.004)
		Stem	43.34	1030 (0.278) to 24 (0.006)
	19	No stem	16.73	1983 (0.534) to 1 (0.000)
		Stem	22.41	1725 (0.465) to 9 (0.002)
	28	No stem	13.13	1933 (0.521) to 0 (0.000)
		Stem	18.58	1453 (0.391) to 0 (0.000)

*Actual numbers for clusters reflect a 60:40 data split to train Naïve Bayes and estimate model accuracy.

Table 4.

Results of Naive Bayes prediction of LDA topics for BA and NI on both validation and test data

Dataset	Number of topics	Stemming?	Naïve Bayes Accuracy (%)	
			Validation data	Test data
BA	7	Stem	57.59	24.32
	8	Stem	57.70	21.78
	9	Stem	56.71	17.34
NI	10	Stem	44.50	15.83
	12	Stem	40.87	8.73
	14	Stem	43.34	11.11

4.3. Main themes identified for clustering and topic modelling

Results described thus far reflected the priority focus on machine learning. Combining machine learning with other techniques was a frequent feature of background research. Table 5 presents a small number of examples of complete student responses for two of the larger clusters for BA data. These were selected as being representative of the predicted cluster based on manually reading the data. However, this represented a potentially significant change of approach from machine learning to human analysis and there are few robust methods to validate how representative these examples may have been. The two clusters outlined in Table 5 accounted for 11% of examples analysed. Similarly, Table 6 presents a sample of responses to two of the larger clusters for NI data. These two clusters accounted for 9.3% of examples analysed.

Table 5.
Examples of responses to larger clusters of BA data

Theme	Selected comments
Work	· Group work
	· Encourage ask questions and encourages team work.
	· To allow us to analyse course work and create work based on what we have learned.
	· Strong emphasis on group work and case-study
Tutorials	· Offers work placement
	· Tutorials with smaller groups
	· Workshops and tutorials
	· Tutorials- better to learn in smaller groups
	· They make sure that each student has a proper understanding of a topic before moving on and helps them if there struggling in tutorials
	· By making the tutorials compulsory in some of my modules, this really forces me to remain engaged

Table 6.
Examples of responses in larger clusters of NI data

Theme	Selected comments
Students	· Offer more helpful services to help struggling students
	· Be more involved with students
	· Maybe organise study groups for students who are struggling and aren't confident enough to ask for help themselves from their peers
	· Listen to students on how they learn individually
	· I believe better engagement between students and lecturers via email or in-person could significantly improve morale amongst students as many of us become frustrated
	· when communication is poor/our worries are unattended to.
Lectures	· More interesting lectures
	· More interactive lectures
	· Encourage people to attend lectures more
	· Record the lectures and put them on Moodle after lectures
	· Not having the lectures so spaced out

An equivalent process of human reading was undertaken for topic modelling. Unlike the data for clustering presented in Tables 5 and 6, topic modelling used the ten most frequently occurring words to describe the “core” theme for each topic. It is acknowledged that stemmed attributes informed allocation of clusters but that variants of individual tokens would remain present in values of the text attribute and, therefore, potentially in the most frequently occurring words in proposed topics. Table

7 illustrates examples of responses to the two largest topics for BA data. These two topics accounted for almost half (49.7%) of all examples categorised. This is in stark contrast to clustering where the two largest clusters contained only circa 10% of examples. Table 8 presents examples from the two largest topics for NI data. These topics included 53.2% of all examples categorized.

5. Analysis of results

Many clusters, as illustrated by Tables 4 and 5, appeared to make intuitive sense. However, analysis of all clusters indicated that some example responses could have been allocated to different clusters. This reflected the fact that documents could belong only to a single cluster whereas responses may refer to multiple issues. However, the distribution of examples to clusters presented a larger problem. As noted, the clusters presented in Tables 4 and 5 represented only 11% and 9.3% of BA and NI responses, respectively. In each case, the largest proposed clusters contained a large proportion of the data with 39.4% of BA examples and 37.2% of NI examples allocated to the largest cluster. These clusters contained, in effect, the examples that had not been allocated to other clusters and did not form coherent themes in themselves but, rather, contained examples where TF-IDF values for all attributes were close to zero. This limitation meant that, while the remaining clusters offered some insights into the data, clustering models effectively did not categorise almost 40% of documents which would significantly limit their usefulness in the business domain. This would particularly be the case when high computational costs are factored in.

Many topics, as illustrated in Tables 6 and 7, featured multiple issues which, in general, appeared to relate quite well to form coherent themes. The top ten most frequently occurring words in each topic provided useful insights into examples contained therein. Unlike clustering, the least intuitive topics accounted for only a small minority of examples at 2.1% of BA documents and 2.7% of NI documents. The largest topics were intuitively feasible and made sense in the business domain. This fact, accompanied by acceptably low computational costs, meant that topic modelling proved to be the most effective method for analysis of these text data, in response to the research question. Some experimental iterations were required to determine optimal numbers of topics but these did not require excessive analyst time.

A number of areas for possible future research were also identified. These include the use of different clustering algorithms to confirm the extent to which difficulties may be associated only with k-means or some other learners. Further research could also seek to categorise the data subset provisionally allocated to the largest clusters, which were found to be uninformative in this research. It may also be informative to disaggregate the data to identify themes or issues that are reported to a greater or lesser extent by particular student cohorts.

6. Conclusion

A structured series of iterations of clustering and topic modelling experiments were undertaken on prepared student responses to prompts on questions about Best Aspects of their educational experience, and what Needs Improvement. Data had been prepared with and without stemming and with the retention and removal of stop words. The best performing k-means clustering models identified allocated a significant minority of examples to single large clusters for both BA data (39.4%) and for NI data (37.2%) which would be problematic in the business domain. However, other clusters were found to be intuitively feasible and should not automatically be discounted. Computational cost for clustering was also regarded as excessive without sufficient benefits to justify that cost. In contrast, topic modelling using Latent Dirichlet Allocation proved to be a computationally efficient method to categorise documents into feasible topics which appeared to be intuitively coherent in the business domain. More than 97% of examples appeared to be appropriately categorised, acknowledging the key distinction, compared to clustering, that examples were assumed to contain multiple topics. Care was needed to determine initial parameter settings and, in particular, to set the number of topics in advance. The use of local maxima for topic coherence values proved effective to inform those choices, whereas perplexity consistently offered apparently optimal values at the maximum number of topics chosen over

Table 7.

Examples of responses for the two largest topics for BA data

Topic ID & proportion	Top 10 words	Selected comments
Topic 4 31.80%	class students classes lectures questions Small small tutorials groups discussion	<ul style="list-style-type: none"> · Small class sizes · Small classes so it's less intimidating to ask questions or speak up in class · Lecturers engage with the students in class by having discussions on topics relevant to the module · Have small groups in classrooms · In my opinion, by incorporating tutorials alongside lectures, it provides an opportunity for students and lecturers to engage and have discussions relating to course topics · Smaller classes allow for discussion and opinions to be said, there's a lot of emphasis on continuous assessment
Topic 1 17.90%	work Group group projects assignments Practical assessment Continuous learning practical	<ul style="list-style-type: none"> · Group work · Lab work · Practical courses with a hands-on approach · Various group work and assignments to keep up to date · Continuous assessment · Group learning Interactive games and quizzes as part of assignments. Presentations of material. Carry out practical course material in class · Having diverse modules and a blend of assignments, real life projects and exams

Table 8.

Examples of responses for the two largest topics for NI data

Topic ID & proportion	Top 10 words	Selected comments
Topic 9 34.9%	lectures class classes tutorials students Make Smaller know interactive activities	<ul style="list-style-type: none"> · More interesting lectures · Have smaller tutorial classes and lectures! · Try and make the lectures and practicals more interactive · More active, hands on lectures, more tutorials where we can work in smaller numbers and have more meaningful discussions, · To have more interaction between lectures and students for all modules · More involvement in lectures · More interactive activities in class
Topic 7, 18.30%	students feedback assignments Provide better lecturers course support Give academic	<ul style="list-style-type: none"> · lectures could be more involved by allowing more time to meet students · More feedback. More emphasis on deadlines. Less readings as is it difficult to balance all academic assignments. · Provide more feedback and direction on future career options · More support from lecturers · A chance to get feedback from academic staff/fellow students on assignments prior to submission. One of our sessions dedicated to this would help. · Give students their exams back when they're corrected

multiple iterations with incrementally increasing numbers of proposed topics. This analysis concurred with background research that human judgement should be used alongside intrinsic measures to determine the optimal initial number of topics.

From experimentation undertaken, topic modelling with stemming proved the most effective method to adopt in future. The key themes contained in responses as identified from analysis were:

Best Aspects

- Smaller classes and tutorials facilitating greater discussions and interactions
- Group / lab work, with practical or “real life” aspects
- Individual engagement with helpful, approachable lecturers
- Listening to students, and various combinations of the aspects listed above

Needs Improvement

- More interaction in lectures; increased smaller group activities
- More feedback
- Greater focus on individual students
- Improved study facilities; more online materials

References

- [1] Williamson, B. (2018). The hidden architecture of higher education: building a big data infrastructure for the ‘smarter university’. *International Journal of Educational Technology in Higher Education*, 15:12
- [2] StudentSurvey.ie (2020). *Irish Survey of Student Engagement National Report 2020*, accessed 10 January 2021, <https://studentsurvey.ie/reports/studentsurveyie-national-report-2020>, p.17
- [3] Rose, J. and Lennerholt, C. (2017). Low Cost Text Mining as a Strategy for Qualitative Researchers. *The Electronic Journal of Business Research Methods*, vol. 15 issue 1, pp. 2-16.
- [4] Gottipati, W., Shankararaman, V. and Lin, J.R. (2018). Text analytics approach to extract courseimprovement suggestions from students’ feedback. *Research and Practice in Technology Enhanced Learning*, vol. 13, no. 6.
- [5] Lee, H., Shimotakahara, R., Fukada, A., Shinbashi, S., & Ogata, S. (2019). Impact of differences in clinical training methods on generic skills development of nursing students: a text mining analysis study. *Heliyon*, vol. 5, no. 3, e01285
- [6] Nitin, G.I., Swapna, G. and Shankararaman, V. (2015). Analysing Educational Comments for Topics and Sentiments: A Text Analytics Approach. *2015 IEEE Frontiers in Education Conference (FIE)*, El Paso, Texas, pp. 1 – 9.
- [7] Santos, C. L., Rita, P. and Guerreiro, J. (2017). Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. *International Journal of Educational Management*, vol. 32, no .3, pp. 431 – 447.
- [8] MacKay, J. (2019). On the Horizon: Making Best Use of Free Text Data with Shareable Text Mining Analyses. *Journal of Perspectives in Applied Academic Practice*, vol. 7, issue 1, pp. 57-64.
- [9] Nikolic, N, Grljevic, O. and Kovacevic, A. (2019). Aspect-based sentiment analysis of reviews in the domain of higher education. *The electronic Library*, vol. 38, no. 1, pp. 44 – 64.
- [10] Boyd-Graber, J., Mimno, D. and Newman, D. (2014). Care and Feeding of Topic models: Problems, Diagnostics, and Improvements. *Handbook of Mixed Membership Models and Its Applications*. 2014, CRC Press.
- [11] Yau, C. K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modelling. *Scientometrics*, 100, pp. 767 – 786.
- [12] Finch, W. H., Hernández Finch, M. E., McIntosh, C. E., & Braun, C. (2018). The Use of

- Topic Modeling with Latent Dirichlet Analysis with Open-Ended Survey Items. *Translational Issues in Psychological Science*, vol. 4, no. 4, pp. 403 – 424.
- [13] Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading Tea leaves: How Humans Interpret Tea Leaves. *Neural Information Processing Systems*. 2009, Vancouver, British Columbia.
- [14] Grebennikov, L. and Shah, M. (2013). Student voice: using qualitative feedback from students to enhance their university experience. *Teaching in Higher Education*, vol. 18, no. 6, pp. 606-618.
- [15] Buenaño-Fernandez, D., González, M., Gil, D., & Luján-Mora, S. (2020). Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modelling Approach. *IEEE Access: Special Section on Advanced Data Mining Methods for Social Computing*, vol. 8, pp. 35318 – 35330.
- [16] Hujala, M., Knutas, A., Hynninen, T., & Arminen, H (2020). Improving the quality of teaching by utilizing written student feedback: A streamlined process. *Computers and Education*, 103965.
- [17] Tsao, H. Y. J., Campbell, C. L., Sands, S., Ferraro, C., Mavrommatis, A., & Lu, S. Q. (2019). A machine-learning based approach to measuring constructs through text analysis. *European Journal of Marketing*, vol. 54, no. 3, pp. 511 – 524.
- [18] Elkan, C. (2003). Using the triangle Inequality to Accelerate k-Means. *Proceedings of the twentieth International Conference on Machine Learning*. Washington, 2003.
- [19] Syed, S. and Spruit, M. (2017). Full Text of Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. *2017 International Conference on Data Science and Advanced Analytics*.