

A Survey on Effect of KPIs in Higher Education based on Text Mining Techniques

Yehia Helmy

Faculty of Commerce & Business
Administration
Helwan University, Cairo, Egypt
yhmhelmy@yahoo.com

Osama Emam

Faculty of Computers and
Artificial Intelligent
Helwan University, Cairo, Egypt
emam_o_e@yahoo.com

Ayman Khedr

Faculty of Computers and
Artificial Intelligent,
Helwan University, Cairo, Egypt
aymankhedr@gmail.com

Mahmoud Bahloul

Faculty of Commerce & Business
Administration, Helwan
University, Cairo, Egypt
mahmoud_bhlol@yahoo.com

Abstract — Higher Education institutions are one of the most important areas that need to use the concept of mining. Even though higher education forms the knowledge foundation for basic professional understanding and development of new capabilities for more information acquisition, Higher Education Institutions (HEIs) need to re-consider their positions, update their curricula according to evolving demands, and provide the services and instructional methods that their client's needs. Key performances indicators are used to assess organizational performance in relation to strategic and operational goals. They are measurable metrics that represent factors that are essential to a university's success. This paper provides a comparison between the most important Text Mining Techniques - such as NLP and Sentiment Analysis - used in Higher Education. It also measures the impact of KPIs in Higher education to evaluate the real state of the universities and to establish the main directions for future action.

Keywords: Text mining, Key Performance Indicators (KPIs), Higher Education, Text Mining Techniques.

1. INTRODUCTION

Higher education expansion has great importance for the country's development and this case should not be overlooked. Performance assessment is also an important issue in higher education. Economic instability, regulatory compliance, technological change, focus on recruitment, public demand for better education threatens higher education today. In order to be able to face these challenges, HEIs need to quickly adapt and adjust what they have done in a mediocre manner, focus on what they do excellent and make quality a critical ingredient [1].

Key performances indicators (KPIs) are the organizations' most detailed goals which direct the activities of the managers to make them achievable. KPIs for higher education are tangible principles that educational institutions use to assess and track their progress towards specific business goals. In addition, these KPIs assist educational institutions in monitoring and evaluating their performance and directing their policy formulation and target setting [2].

With the increasing of information and recourse texts, text mining has become an important and more favorite domain to maintain and display the main purpose of textual knowledge. There is a drastic increase in the amount of text produced every day. This terrific extent of basically unstructured textual content cannot be merely processed and perceive this tremendous volume of data. Consequently, to discover helpful patterns, efficient and effective techniques and algorithms are needed [3].

Text mining is the method of searching for or extracting useful information from textual content. It is an interesting area of research as it aims to learn from unstructured documents. TM is a vital and powerful field of computer science research which aims to solve data mining, machine learning, information extraction, natural language processing, information retrieval, knowledge management, and classification issues [4].

The scope of this research paper presents literature on the use of Text mining and KPIs in Education. This paper implemented two phases to classify and extract related articles

and report literature-based results. The first stage is identifying and extracting the articles that related to Text mining in Education. The second stage is identifying and extracting the articles that related to KPIs in Education.

2. TEXT MINING IN HIGHER EDUCATION

Learning is the core business of the universities. Cognitive reasoning is important for learning and developing new knowledge that is necessary for higher education. To search the web for related content, in [5] used text mining tool to extract relevant terms and keywords from the writings of the students. Recommendation systems have been proposed as a way to help users find items of interest that aim to minimize this challenge by recognizing the desires of users according to the things they analyze. Results of the experiment involving 34 graduate and undergraduate students have shown that they have a positive perception of their use of the tool in the activity growth. Students presented the instrument with an average score of 8.24 in a scale of 0 to 10, and a score of 8.03 for the learning activity based on inquiry.

In [6] aimed to reduce the plagiaristic conduct of learners in online assignments by offering automated feedback based on an analysis of text mining. Text similarity analysis was conducted on 4,268 reflective texts (max. 500 characters) written by the participating university students (N=59) at the middle and end of the semester on concepts they had learned in the course of computer science. The findings of the analysis showed the ratio of plagiarized posts decreased by an average of 21.07 percent, 83 percent of students had lower plagiarized post ratios. After reviews, the number of students who did not commit plagiarism increased by 42.37%.

The paper [7] examined the degree to which the lexical properties of the essays of students can inform their vocabulary knowledge of stealth evaluations. Examine the effectiveness of NLP techniques to inform stealth evaluations of this knowledge. They used calculated indices using the natural language processing tool to predict the performance of students on a measure of vocabulary knowledge. The results showed that two of the linguistic indices could account for 44

percent of the variance in the vocabulary knowledge scores of the college students.

In [8] study interested in identifying students who are confused as to their answers. Results suggest the following, Words alone are not adequate to accurately predict the results at item level. The DALITE platform data has the ability to provide teachers with a novel forum for formative evaluation that forecasts progress and learning for students.

The study [9] proposed framework aims to engage language learners by creating multiple-choice questions that use different inference measures over several sentences, namely the resolution of coreferences and paraphrases. The aim of the question generation system is to improve the ability of language learners to read, in particular students learning English as a Second Language (ESL). This study have provided two different methods that produce multiple-sentence questions automatically. Experiments have shown that questions created by both approaches require more inferential steps than conventional single-sentence approach questions. The first approach, in particular, outperforms the baseline procedure on average by 0.60 in terms of the number of inference measures, while maintaining comparable grammatical correctness and the presence of the response.

The research [10] designed a series of Natural Language Processing (NLP) algorithms to help triple concept-relation-concept extraction to shape maps for concept. In order to rank the triples according to their value, structural and graph-based features are used. The results showing a statistically significant correlation ($r_s > 0.455$) between the auto-generated concept maps and the judgment of human experts. Auto-generated concept maps were graded as a pedagogical tool by academics on evaluation criteria such as coverage, accuracy and appropriateness from 'good' to 'very good'.

The paper [11] concentrated on how we construct the Sentiment Text Recognition module. This takes the views of the students on the machine programming exercises, and then the module will be rendered Sentiment Analysis to decide whether the polarity of the exercise is positive or negative (like or not like). Finally, teachers will be able to assess the opinion of the students and decide the consistency of the programming exercises.

The result got an accuracy of 80.75 percent which means that in 80.75 percent of cases it got a correct prediction about text polarity. The accuracy test 0.81 tells us how many related tweets we found are true-positive.

The research [12] explained the use of textual data analytics focused on Machine Learning and Natural Language Processing to understand the word choices in U.S.-based engineering colleges' mission statements. With 59 engineering colleges in the United States, it used a Word2Vec technique to represent the terms from mission statement. The purpose of this paper is to visually map the word choices between private and public engineering colleges in the USA by using the embedding word2vec they found that the word choices used by the public and private colleges overlapped.

In [13] has been explained by their lecturers to the students. During the lecture, the students were asked to provide feedback via Twitter, depending on the preference of the lecturer. They had to tweet an anonymous Twitter account or they could use their own accounts. The study developed the SA-E system to analyze real-time feedback from students through social media and also present the evaluation of this system with lecturers and students in real settings. The student responses showed that 85 percent of students received feedback via Twitter, showing their willingness to provide real-time feedback. This indicates that, by informing the students how their feedback is used, the "feedback loop" needs to close. The results show that students were highly involved in providing feedback, and the opinions of the students about the system were neutral and dislike.

In this work [14] presented a course-adaptive recommender system that helps programming training instructors pick the most appropriate learning materials. The recommender system uses program examples prepared by the instructor to deduce the envisaged structure of a specific course and recommends learning content items adapting to the instructor's intentions. This study also present a study that used data sets collected from different courses to assess the quality of recommendations. This paper collected three data sets from the University of Pittsburgh, the USA and Helsinki University, Finland to evaluate their recommendation method. In curriculum collection, each data set encapsulated teacher expectations. The Content Wizard outperforms the baseline for all datasets.

TABLE 1. SUMMARY DESCRIPTION OF TEXT MINING TECHNIQUES IN HIGHER EDUCATION

Ref	Year	Objective(s)	Text mining Technique	Benefits	Limitations
[5]	2014	This study aims to allow students to use mobile devices to create and share questions and answers by using the SMILE project (Stanford Mobile Inquiry Based Learning Environment).	NLP	This research has followed a hybrid recommendation approach in an attempt to take advantage of both the possibility of defining access trends for educational material and filtering them according to their content.	Needs some fine tuning to optimal sizes. For efficient record matching methods more effort is needed. More effort is needed for accurately identify languages relationships.
[6]	2015	The objective of this study address the issue of digital plagiarism in the context of educational data mining and	Sentence similarity	This study addressed problems with digital plagiarism commonly used by students such as plagiarism committed online and	Needs more effort for can improve performance of algorithm for detecting text similarities to get

		whether the automated feedback received by text mining was effective in preventing this serious problem which is hindering online learning		students copying each other's work.	faster results
[7]	2015	This study aims to examine whether the students' essay lexical properties can accurately model their scores on a standardized measure of vocabulary awareness to provide more individualized tutoring to student users.	NLP	The use of NLP techniques in computer-based learning environments to help researchers and program developers create covert tests and student models. Investigate the effectiveness of NLP techniques to notify covert of vocabulary knowledge appraisals.	Further effort is needed to clearly analyze how these vocabulary awareness stealth tests will shift throughout the training and how they will help to notify student model changes consistently.
[8]	2016	The first aim of this study is to decide whether the terms used by students in their self-explanations can predict their performance on the related multiple-choice object, or even expose their confusion over the concept being studied. The second aim is to assess whether a student's collection of words using The Distributed Active Learning Integrated Technology Environment (DALITE) during a semester will predict their end-of-semester learning outcomes.	Classification	This study analyzes the DALITE platform data and has no dialog agents, but has students interacting asynchronously with the self-explanations of each other. This study based work on developing tools to engage students, even outside the classroom, in active learning.	This paper does not expressly produce any results from the first set of experiments, except in saying that none of our statistical models and wasn't able to achieve a prediction accuracy above baseline. The most important facet that has not yet been studied lies in the patterns in student preferences and crowdsourcing.
[9]	2016	The propose of this work is a question generation approach that engages learners by using different inference measures through several sentences, including paraphrase identification.	NLP	Introduce a novel approach to automated question generation that enhances prior work from both a technical and evaluation perspective.	Need to explore more strategies for the generation of question-adaptable domains.
[10]	2017	This work explores the efficacy of automated methods in extracting concept maps from lecture slides and the appropriateness of self-generated concept maps as a pedagogical tool.	NLP	Research helps bridge the gap between linearity in teaching materials and the need to create integrated teaching resources network models. Auto-generated concept maps can	Architecture is not predicated on external resources. The availability of slides of lecture are adequately minable.

				be used as a positive alternative to expert concept maps manually.	
[11]	2017	This study introduces a Sentiment Analysis module implementation that operates in an Affective Intelligent Tutoring System.	Sentence similarity	This work will allow the system to evaluate student feedback using sentiment analysis techniques and to provide useful information to teacher course administrators to enhance teaching material.	This study exclude multimodal affect recognition and identification of emotion by text, voice, and EEG devices.
[12]	2017	The aim of this study is to understand the key similarities and differences between the words used in the two groups ' mission statements: public colleges and private engineering colleges.	NLP	This research represented as in the form of a diagram mapping the vector space model for word usage and a full vocabulary of relevant words from the statements being analyzed.	The lack of inclusivity, as the application was limited to only one college and also not taking into account many other aspects of data analysis such as individual institutional characteristics.
[13]	2017	The objective of this paper integrated the sentiment analysis models and the visualisations into a PC-based system to be used in real lectures, and the opinions of the students and lecturers were collected and analysed.	Sentence similarity	Evaluation of models of sentiment analysis using machine learning methods to provide polarity feedback distribution like positive / negative / neutral and also presence in reviews of specific learning-related emotions	Feedback loop needs to close by informing the students how their feedback is used. The lecturers made no changes according to feedback from the students.
[14]	2018	The research extends the functionality of an introductory programming courses tool which supports instructors in selecting smart learning content for students. Smart material may be chosen from a large pool of parameterized problems or annotated examples by instructors.	Clustering	Assist instructors in the authoring process through a content recommender that suggests learning activities that are most appropriate to instructors' intended model of the course.	This work with instructors has shown that the assistance offered by the current course authoring tool is not enough. The instructors need to examine a large number of problems and examples carefully, this is a time-consuming and error-prone process.

3. KPIs in Higher Education

This work [15] provided a practical framework, model and guidelines which institutions of higher education can use to assess and improve their performance, and better prepare students to work effectively in society. The context, input, process, product (CIPP) model is a fitting macro-evaluation model for institutions of higher education. With nine perspectives on metrics and measures conceptualized, explained and implemented in this paper, fulfills this gap through changing institutions of higher education from a simple examination of past performance to designing future strategies. Nine perspectives were developed and explored in performance assessment using the CIPP model and its implementation in higher education institutions in the USA largest university, by student enrollment.

This research [16] aimed to establish qualitatively the essential Culture of Excellence (CoE) variables for PHLI employees across Malaysia. The process of this research started with the identification of the CoE determinants, followed by the development of the research instrument and the pilot test. This study provides the solution needed to use the Culture of Excellence (CoE) as a resource by the Top Management of Public Higher Learning Institutions (PHLI). The findings revealed five (5) new items have been identified which include integrity and accountability, positive traits, freedom from political elements, The first two (2) items related to the individual staff and their control and the other three (3) items are external factors.

In [17] discussed e Canadian higher education system is regulated provincially, with policy and funding decisions relevant to provincial jurisdictional higher

education. This study describes the perception of the participating universities' major informants about the effectiveness and effectiveness of existing KPIs. This study looked at 12 key informants' perceptions at 11 participating universities. In the reporting of key performance indicators (KPIs) over the years, the relevant documents and existing data were analyzed to understanding of the effect of the KPIs on these universities' results. For this study a sequential mixed method and exploratory descriptive design were used.

This work [18] tried to develop a framework for KPIs that takes into account student, faculty and institutional development which faculties have an enormous task to finalize every student as an individual entity on the end result. Implementing such a performance management system (PMS) will help create the ranking of the faculty, meet the needs of the stakeholder and will also help sustain the ever-increasing competition.

This paper [19] explored the concept of KPI from the viewpoint of software metrology, assessment methods, organizational contexts, and professional views through developed a good KPI model that meets 59 quality attributes of the model described in this paper. Completing these qualitative attributes makes the KPI compatible with standard software engineering comparison models, the organizational context of the KPI, the quality of the KPI as a data point and the KPI's capacity to induce organizational action. The results show the model is coordinating effective KPI development practices and stakeholders can easily use this to improve the quality of the KPIs or to reduce the number of KPIs.

The paper [20] presented a Riga Technical University case of use, which exemplifies the use of Capability-driven-development (CDD) in the field of IS governance. CDD is a capacity management strategy that maps the skills of the organization to IT solutions, captures history surrounding the implementation of requirements and the subsequent requirement changes that ensure the attainment of goals and previously identified KPIs in contextual variable situations. The CDD allows the separation of data integration and interpretation from actions needed to respond to the current contextual situation. Also CDD allows the IS governance solution to be more maintainable and traceable.

In [21] Emphasizing on the Demand Response (DR) concept which provides customers with an ability to play an important role in the operation of the electric grid by reduce or change their electricity consumption during peak periods in response to time-based tariffs or other financial incentives. The calculation of the KPIs for the Romanian demonstration site for the evaluation of the DR through different scenarios were identified and the DR events were applied over the entire evaluation period, between January 2018 and December 2018. Three scenarios have been developed, primarily aimed at temporarily reducing TUCN peak power demand by shifting / rescheduling the use of different equipment in line with national daily peak power demand and local Technical University of Cluj-Napoca (TUCN) peak power demand.

TABLE 2. SUMMARY DESCRIPTION OF KPIs IN HIGHER EDUCATION

Ref	Year	Objective(s)	Benefits	Limitations
[15]	2016	The aim of this research is to establish a structure for performance evaluation of higher education institutions using the context, input, process, product (CIPP) model.	Provide a good basis for measurement through assessing the performance of a higher education institution from input-process output view.	The model does not adequately address measurements and comparisons.
[16]	2015	The purpose of this study is provides recommended a number of key performance indicators which might better serve all stakeholders, that applied on Ontario Student Assistance Programme	Awarding the performance funding at the system and institutional levels through aggregation. This is far from the position of plan management decision making, which is usually at the (dean) stage of the programme.	Need to indicators to address performance of decision-making whether at the institutional or programme level
[17]	2015	This research aims to identify, point out and evaluate some of the Culture of Excellence (CoE) dimensions.	Increasing employee's productivity through implementing the culture of excellence in their respective organizations. Prompt the Culture of Excellence (CoE) among employees that will motivate and increase their loyalty and satisfaction.	No initial structural model was submitted to test the results obtained from the questionnaires.

[18]	2016	The paper aims, firstly, the categorization of performance indicators in broad areas, and secondly to define key performance indicators (KPIs) for evaluating faculty performance.	The implementing performance management system (PMS) will help Institutes to raise their standards of quality.	The KPIs in this study doesn't provide guidance on action to be taken and there is no approach applied to check this metrics.
[19]	2016	The paper's goal is to create, document and evaluate a quality model for KPIs and addressing the research question of what characterizes a good KPI? Also to find a KPI to be "healthy or good" when it can be applied and helps the company within achieving its strategic objectives.	Using metrics for medium to large software development organizations of our time, which often depend on quantitative information in their product and process monitoring.	The KPIs in this study have not been tested operationally. More data is needed to enable statistical analysis of whether the quality attributes correlate in practice.
[20]	2018	The goal of this work is to model the capabilities of IS security governance in Riga Technical University (RTU) that will act as a foundation for further improving the existing approach.	Use of CDD has allowed data integration and analysis to be segregated from the reaction logic, which will make IS governance system more efficient and traceable.	This approach based on only one assumption that IT security staff is unable to control most of the user devices. Each KPI was previously defined while the KPI targets and baseline method not mentioned. This approach doesn't predict KPIs values and status.
[21]	2019	This paper illustrates the pilot site evaluation, concentrating on the explanation of the KPI's calculation method and the findings obtained at the Swimming Pool Complex.	Evaluation and computation of the KPIs for each TUCN location involved in DR events.	The KPIs proposed need to be calibrated regularly. Need for using this strategy for the other sectors.

4. CONCLUSION

At last we conclude that, Text Mining is typically used to achieve rapid results and carrying out work in number of areas of use. Text Mining may be classified as Text Clustering, NLP and other techniques based on the respective areas of use. This survey tries to provide a detailed view of four Text Mining Techniques in Higher Education Institutions and the KPIs' role in developing education. In the near future, a framework will be proposed as an improvement over the limitations suggested in the analyzed techniques of the study.

5. REFERENCES

- [1] Briqa'an, A. M. A., and A. A. Alqurashi. "Governance and role of universities in the face of challenges." th International Scientific Conference. 2012.
- [2] Kairuz, Therese, et al. "Consequences of KPIs and performance management in higher education." International Journal of Educational Management 30.6 (2016): 881-893.
- [3] Lin, Hsueh-Yuan, and Sheng-Yuan Yang. "A cloud-based energy data mining information agent system based on big data analysis technology." Microelectronics Reliability 97 (2019): 66-78.
- [4] Thu Vo Anh, AU - Wanbing Shi. "Trend of higher education abroad policy analysis based on regression analysis and data mining." 2017 International Conference on Sports, Arts, Education and Management Engineering (SAEME 2017).
- [5] Acosta, Otavio Costa, Patricia Alejandra Behar, and Eliseo Berni Reategui. "Content recommendation in an inquiry-based learning environment." 2014 IEEE Frontiers in Education Conference (FIE) Proceedings. IEEE, 2014.
- [6] Akçapınar, Gökhan. "How automated feedback through text mining changes plagiaristic behavior in online assignments." Computers & Education 87 (2015): 123-130.
- [7] Allen, Laura K., and Danielle S. McNamara. "You Are Your Words: Modeling Students' Vocabulary Knowledge with Natural Language Processing Tools." International Educational Data Mining Society (2015).
- [8] Bhatnagar, Sameer, et al. "Text Classification of Student Self-Explanations in College Physics Questions." EDM. 2016.
- [9] Araki, Jun, et al. "Generating questions and multiple-choice answers using semantic analysis of

- texts." Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016.
- [10] Atapattu, Thushari, Katrina Falkner, and Nickolas Falkner. "A comprehensive text analysis of lecture slides to generate concept maps." *Computers & Education* 115 (2017): 96-113.
- [11] Barrón-Estrada, María Lucía, et al. "Sentiment analysis in an affective intelligent tutoring system." 2017 IEEE 17th international conference on advanced learning technologies (ICALT). IEEE, 2017.
- [12] Bhaduri, Sreyoshi, and Tamoghna Roy. "A word-space visualization approach to study college of engineering mission statements." 2017 IEEE Frontiers in Education Conference (FIE). IEEE, 2017.
- [13] Altrabsheh, Nabeela, et al. "Evaluation of the SA-E system for analysis of students' real-time feedback." 2017 IEEE 17th international conference on advanced learning technologies (ICALT). IEEE, 2017.
- [14] Chau, Hung, Jordan Barria-Pineda, and Peter Brusilovsky. "Learning content recommender system for instructors of programming courses." *International Conference on Artificial Intelligence in Education*. Springer, Cham, 2018.
- [15] Chinta, Ravi, Mansureh Kebritchi, and Janelle Ellias. "A conceptual framework for evaluating higher education institutions." *International Journal of Educational Management* (2016).
- [16] Chan, Vivian. "Implications of key performance indicator issues in Ontario universities explored." *Journal of Higher Education Policy and Management* 37.1 (2015): 41-51.
- [17] Rahman, Nasriman Abdul, Boo Ho Voon, and Abdullah Firdaus. "Identifying the dimensions for culture of excellence in higher education." *Procedia-Social and Behavioral Sciences* 224 (2016): 84-92.
- [18] Joshi, Sangeeta M., et al. "Developing Key Performance Indicators framework for evaluating performance of engineering faculty." 2016 IEEE Eighth International Conference on Technology for Education (T4E). IEEE, 2016.
- [19] Staron, Mirosław, et al. "A key performance indicator quality model and its industrial evaluation." 2016 Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA). IEEE, 2016.
- [20] Minkevics, Vladislavs, and Janis Kampars. "IS Security Governance Capability Design for Higher Education Organization." 2018 59th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS). IEEE, 2018.
- [21] Cretu, Mihaela, et al. "Key Performance Indicators (KPIs) for the Evaluation of the Demand Response in the Technical University of Cluj-Napoca Buildings." 2019 8th International Conference on Modern Power Systems (MPS). IEEE, 2019.

IJSER