International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:12, No:6, 2018

A Dataset of Program Educational Objectives Mapped to ABET Outcomes: Data Cleansing, Exploratory Data Analysis and Modeling

Addin Osman, Anwar Ali Yahya, Mohammed Basit Kamal

Abstract**—**Datasets or collections are becoming important assets by themselves and now they can be accepted as a primary intellectual output of a research. The quality and usage of the datasets depend mainly on the context under which they have been collected, processed, analyzed, validated, and interpreted. This paper aims to present a collection of program educational objectives mapped to student's outcomes collected from self-study reports prepared by 32 engineering programs accredited by ABET. The manual mapping (classification) of this data is a notoriously tedious, time consuming process. In addition, it requires experts in the area, which are mostly not available. It has been shown the operational settings under which the collection has been produced. The collection has been cleansed, preprocessed, some features have been selected and preliminary exploratory data analysis has been performed so as to illustrate the properties and usefulness of the collection. At the end, the collection has been benchmarked using nine of the most widely used supervised multiclass classification techniques (Binary Relevance, Label Powerset, Classifier Chains, Pruned Sets, Random k-label sets, Ensemble of Classifier Chains, Ensemble of Pruned Sets, Multi-Label k-Nearest Neighbors and Back-Propagation Multi-Label Learning). The techniques have been compared to each other using five well-known measurements (Accuracy, Hamming Loss, Micro-F, Macro-F, and Macro-F). The Ensemble of Classifier Chains and Ensemble of Pruned Sets have achieved encouraging performance compared to other experimented multi-label classification methods. The Classifier Chains method has shown the worst performance. To recap, the benchmark has achieved promising results by utilizing preliminary exploratory data analysis performed on the collection, proposing new trends for research and providing a baseline for future studies.

Keywords**—**Benchmark collection, program educational objectives, student outcomes, ABET, Accreditation, machine learning, supervised multiclass classification, text mining.

I. Introduction

THE Accreditation Board for Engineering and Technology (ABET) paved the way for the accreditation process in the United States since 1932. It has been widely accepted and recognized as a leading accreditor of colleges and university programs in applied science, computing, engineering and technology in the US and worldwide. Through its accreditation activities and dedication, ABET has accredited more than 2,700 programs at over 550 colleges and universities nationwide [13].

ABET delegates to the Engineering Accreditation Commission (EAC) the responsibility for evaluating and taking accreditation actions on baccalaureate programs in engineering. The EAC depends on experts, volunteers, visiting teams to perform the crucially important and fundamental evaluation that is the basis of the EAC/ABET accreditation process. The engineering programs applying for ABET accreditation must submit to ABET their Self-Study Reports (SSRs). SSRs are the primary documents, which programs use to demonstrate their compliance with all applicable ABET criteria and policies. The SSR is the foundation for the review team's judgment of whether the program meets the ABET criteria for accreditation. It addresses all paths to completion of the degree, all methods of instructional delivery used for the program, and all remote location offerings [13], [14].

All programs applying for ABET accreditation must have documented program educational objectives (PEOs) that are consistent with the mission of the institutions and the needs of the programs' various constituencies. ABET defines PEOs as the expected accomplishments of graduates during the first few years after graduation.

An efficient periodic review and revision of the PEOs must be documented by the programs. Consequently, the programs must have documented student outcomes (SOs), which specify the desirable abilities, attitudes and knowledge that will be attained by the students immediately after graduation [13]-[15]. Most if not all PEOs are manually mapped to the ABET SOs, which are adopted by many Engineering programs. They represent one of the most important parts of SSRs. PEOs are significant as they represent the eventual mean to judge the quality of a program. They are directly mapped to the SOs and indirectly to all course learning outcomes (CLOs) in the program. Each and every CLO must be directly mapped, least to one SO.

This research is an endeavor to provide a dataset collection of mapping PEOs to SOs and to shed light on a rich new research area (accreditation) for the researchers and to perform preliminary exploratory data analysis on this dataset to uncover its structure.

The rest of the paper is organized as follows. In Section II, we start by presenting the importance of the dataset collection. In Section III, the collection has been presented containing the operational setting under which the collection has been produced, the content of the collection, and the quality of the collection. Section IV presents the data cleansing and

Addin Osman and Mohammed Basit Kamal are with the College of Computer Science and Information Systems, Najran University, Saudi Arabia (e-mail: aomaddin@nu.edu.sa, mbkamal@nu.edu.sa).

Anwar Ali Yahya is with the College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia and Faculty of Computer Science and Information Systems, Thamar University, Thamar, Yemen (e-mail: aaesmail@nu.edu.sa).

exploratory data analysis implemented on the collection. In Section V, the benchmarking of the collection to different multi-class classification has been presented. In Section VI, the results of the benchmark have been presented. Lastly, in Section VII, a summary of the work and future directions of the research has been presented.

II. Importance of the Data

In [16], it has been mentioned that much of the data collections represent the base for the scientific research and they are changing the face of the world. Scientific data collections are at least intermediate results in many scientific research projects. An interesting situation is that data collections are becoming more significant themselves and can sometimes be considered as the primary intellectual output of the research.

Data collections do not exist in a vacuum. The quality and usefulness of the collections depend on the context on how they are collected, processed, analyzed, validated, and interpreted; thereafter, it can be used to lead to a scientific publication. The process leading to a scientific publication can be described by the following basic scientific workflow shown in Fig. 1 [16].

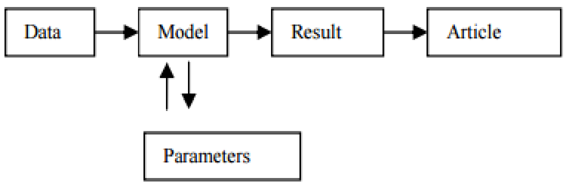

Fig.1 The Process Leading to Scientific Publication [16]

The research interests in supervised and unsupervised techniques has been tremendously increasing in machine learning, information retrieval, natural language processing, computational linguistic, and data mining. Text categorization, clustering and association rules are significant application areas for machine learning and mainly depend on the availability of the collections [1], [8]-[12]. Generally speaking, the academic programs planning for accreditation or reaccreditation prepare their text collections depending on human experts to assign class-labels from a predefined set. The accreditation bodies and academic programs also depend on human experts to check, read and comprehend these data collections, specifically in the case of PEOs. First, they have to check the satisfaction of the PEOs to the specification of learning objectives (action verbs, condition, and standard), and second, to check the veracity of the mapping to the appropriate SOs. It is inevitable that a large amount of manual human effort, resources and money are required [17]-[19].

The test collections, especially unstructured collections in accreditation, are suffering from many weaknesses such as: lack of complete text, peculiar textual properties, scattered test collections, and/or limited availability of datasets. These difficulties are worsened by the shortage of documentation providing the methods used in producing these collections and on the nature of their classification systems. The researchers interested in supervided and unsupervised learning techniques in this area are facing serious problems, because of the shortage of well analyzed collections and corpora. They have often been compelled to impose their own assumptions on these collections [11], [20], [21].

III. Data Collection

In this research, we have considered the data of the SSRs that are produced in an operational settings at ABET accredited engineering programs under procedures managed and operated by specialists and experts in the area. The data were seldom used in research, and only recently has its use for this purpose been contemplated.

We have sent requests to many of the ABET accredited programs in different countries requesting their SSRs. Nevertheless, we have not received any response from any program. At the end we have to collect data from some of the SSRs available online.

A. Documents

The SSRs are one of the largest international sources of data that are significant in accreditation. Hundreds of programs in engineering have been accredited by ABET worldwide. The SSRs are produced in English language and many of them distributed and made available online in PDF and word formats.

The data of the mapping of PEOs to ABET SOs were drawn from those SSRs. The data consisted of only English language objectives produced by ABET accredited programs were considered. The data collection has been formatted in XML.1, Excel and ARFF formats. The preparation of the dataset involved substantial verification and validation of the content, attempts to remove spurious or duplicated objectives, fulfilling the objectives and outcomes format, etc.

B. Class-Labels (Student Outcomes)

All ABET accredited programs must have documented PEOs, which are compatible with the institution's mission and the requirements of the diverse constituencies of the program. ABET defines PEOs as the expected accomplishments of graduates during the first few years after graduation. The periodic review and revision of the PEOs must be documented. In addition to that, SOs must clearly present what students will know and be able to do by the time of graduation. These are related to the knowledge, skills, and behaviors that students develop while advancing in the program. The PEOs were manually mapped to the 11 ABET SOs, which are considered as the class labels. One PEO can be mapped to more than one SO (supervised multiclass classification). SOs are outcomes (a) through (k) plus any additional outcomes that may be articulated by the program [14]. In our case, we have considered only the ABET SOs, which are found in [11], [15]:

1. Policies of Developing PEOs

Developing POE policies specify certain requirements to be

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:12, No:6, 2018

satisfied while developing PEOs. In this section, some of the well-known policies employed in developing PEOs are presented. The PEOs of a program should clearly specify the intent of the program and what it is expected to achieve. There should be clearly no any room for misinterpretation and misunderstanding. Every PEO must start with a relevant action verb selected from the Bloom's Taxonomy that specify and definite observable and required behaviors [17]-[19].

POEs should also specify or state what learners are expected to perform or be able to do some years (3 to 5 years) after completion of the program. A very important point to be noted is that each and every learner completes a program should be able perform innovation in and enhance his profession. Beside that the PEOs are prefered to be **SMART**, which pointing to:

- **S**pecific: The PEO must be distinct about when, how, what and where the situation of students will be changed.
- **M**easurable: The goals and benefits must be quantifiable (can be measured).
- **A**chievable: The objectives must be attainable, that means, they can be done within an specific time framework with the available resources.
- **R**ealistic: It can extend the abilities reflected on the objectives but remain achievable.
- **T**ime bound: It must specify clearly the time by which the objectives must be achieved (e.g. upon successful completion of the XYZ program).

IV. Data Cleansing and EDA

As mentioned previously, the data used in this research has been collected from 32 ABET SSRs available online. Data collection is a tedious, complicated and difficult task. Generally, organizations spend huge amounts of money, manpower and resources on data collection so as to build better models in a perfect world. The data collection process is error prone, and in big organizations it involves many steps. Therefore, it is a good practice to interpose, capture and fix data errors as much as possible at the early stages before engaging on data visualization and modeling [26].

A. Data Cleansing

The data used in this research have also been passed through different cleansing processes, and then transformed into a usable form. Some libraries of Natural Language Processing in Python have been utilized for the data cleansing and preliminary Exploratory Data Analysis (EDA) [28], [30]. At the beginning, the vocabularies of the collection of all PEOs have been computed, and then, all items that occur in an existing wordlist have been removed. Only the uncommon or misspelled words have been left. As a result, no uncommon or misspelled words have been found.

Some simple statistics have been performed on the dataset. The average length of words (tokens) in the PEOs collection is 6.20. The number of all tokens is 8118 and the most 50 frequent words have been derived from the collection. They are shown below together with their frequencies: (and, 591), (to, 340), (in, 307), (the, 256), (of, 245), (engineering, 194), (will, 132), (professional, 117), (their, 112), (a, 109), (or, 102), (graduates, 94), (skills, 81), (be, 65), (ability, 64), (as,62), (an, 59), (with, 56), (design, 53), (problems, 52), (for, 50), (students, 49), (knowledge, 49), (learning, 48), (effectively, 47), (technical, 45), (systems, 45), (our, 43), (work, 43), (graduate, 42), (development, 42), (apply, 41), (have, 39), (through, 38), (demonstrate, 38), (ethical, 35), (leadership, 33), (program, 32), (environmental, 31),(pursue, 30), (practice, 28), (education, 28), (career, 27), (technology, 27), (life, 27), (lifelong, 26), (), 26), ((, 26), (on, 25), (communicate, 25).

Fig. 2 shows the cumulative frequency plot for 50 most frequently words in the collection. These account for nearly half of the tokens.

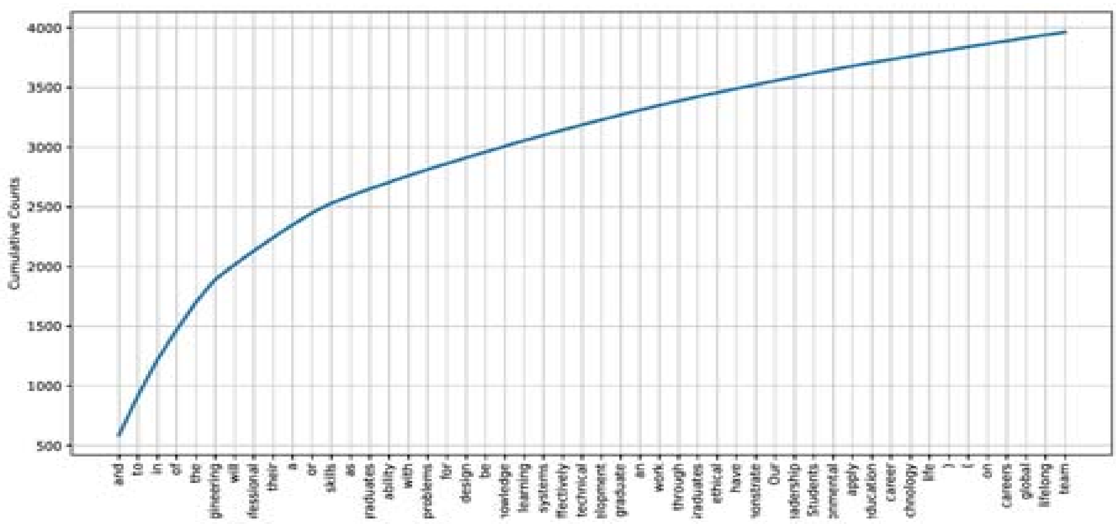

Fig. 2 Cumulative Frequency Plot for 50 Most Frequently Words in the Corpus: These Account for Nearly Half of the Tokens

The tokens of length greater than or equal to 15 are:

professionalism, competitiveness, multidisciplinary, collaboratively, instrumentation, internationally, microprocessors, interdisciplinary, telecommunications, responsibilities, electromechanical, interrelationships, entrepreneurial, accomplishments, microcontrollers.

Here are all tokens from the corpus that are longer than seven characters, which also occur more than seven times (87 words were found):

Graduates, Students, activities, advanced, analysis, applying, appreciation, appropriate, awareness, changing, commitment, communicate, communication, community, components, computer, continue, continuing, continuous, critical, demonstrate, development, disciplinary, disciplines, economic, education, effective, effectively, electrical, electronic, engineering, engineers, environment, environmental, experiments, function, fundamental, graduate, graduates, identify, improvement, including, industrial, industry, information, knowledge, leadership, learning, licensure, lifelong, management, mathematics, mechanical, multidisciplinary, necessary, organizations, participate, personal, practical, practice, practices, prepared, principles, problems, processes, productive, profession, professional, professionally, programs, research, resources, responsibilities, responsibility, societal, software, solutions, students, successful, successfully, teamwork, technical, techniques, technology, thinking, training, understanding

The process of searching for words with long lengths helped in detecting and fixing many two concatenated words (e.g. "*willdemonstrate*" and "*lifelonglearning*"). The concatenation of words resulted from combining *PEOs* from different sources and transforming them from *Exce*l format to text format.

The process of checking words with long lengths and with higher frequencies has helped us to succeed in automatically extracting words with a miss-typified text. Well, these very long words are often hapaxes (i.e., unique) and perhaps it would be better to find frequently occurring long words. This seems promising since it eliminates frequent short words (e.g., *the*) and infrequent long words promising since it eliminates frequent short words (e.g., *the*) and infrequent long words (e.g., *electromechanical*).

A collocation is referring to a series of words that unusually often appear jointly. In our case, we would like to attract attention in a situation where rare words are important. Specifically, we would like to find bigrams that frequently occur together more often than we would anticipate based on the frequency of singular words. The collocations that we have extracted from the dataset collection are:

lifelong learning; lifelong; long learning; civil engineering; communicate effectively; ene program; leadership roles; professional development; related fields; communication skills; multidisciplinary teams; problem solving; successful careers; engineering problems; critical thinking; continuing education; function effectively; graduate studies; mechanical systems; self-improvement

B. Exploratory Data Analysis

EDA is a process for summarizing, visualizing, and becoming knowledgeable about the significant characteristics of a dataset. It helps to ensure the readiness of selecting and using the most appropriate Machine Learning (ML) techniques (e.g. supervised or unsupervised models) and selecting the best features that potentially can be used for these techniques. Implementing ML techniques by entirely skipping EDA is a big mistake with many implications. These implications include generating inaccurate models, generating accurate models but on wrong data, not creating the right types of variables in data preparation, or using resources inefficiently because of realizing only after generating models that perhaps the data is skewed, or has outliers, or has too many missing values, or finding that some values are inconsistent [27].

At an upper level, EDA is the practice of employing visual and quantitative methods to comprehend and recap a dataset without making any assumptions about its content. It is a critical process needed to develop an appropriate model for the problem at hand and to correctly interpret its results. It is valuable to be certain that the results that will be obtained by ML techniques are valid, correctly interpreted and applicable to the desired context. In addition, it helps in ensuring that the potential value of the output will be maximized [27]-[29].

The EDA process for the dataset used in this research has been started with driving a list of frequencies of token lengths in the corpus. The result is a distribution containing thousands items, each of which is a number corresponding to a word token in the collection (Fig. 3). But there are only 18 distinct items being counted, the numbers 1 through 18, because there are only 18 different word lengths (the words in which their lengths are ranging from 1 to 18 characters and not more than that). From this, it can be concluded that the most frequent word length is 2, and that words of length 2 account for roughly 1280 (or 16%) of the words making up the corpus.

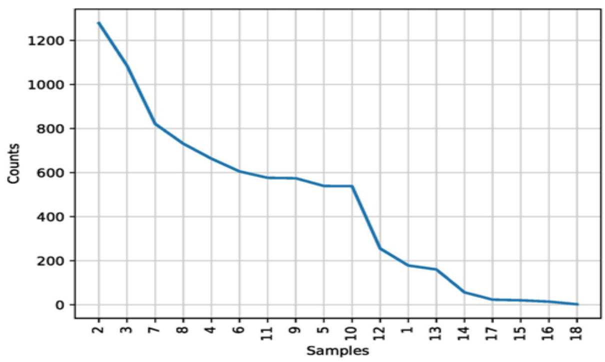

Fig. 3 The Frequencies of the Lengths of the Tokens in the Corpus

The second EDA process performed in the research is the frequencies of the first and last letters of the tokens in the corpus. Fig. 4 shows the distribution of the frequencies of the first and last letters of the tokens in the collection.

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:12, No:6, 2018

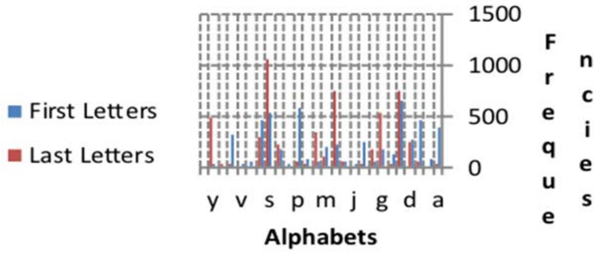

Fig. 4 Frequencies of First and Last Letters in the Tokens

Fig. 4 presents some correlation between the frequencies of the first and last letters in the corpus. The correlation seems not very strong; this might be attributed to not eliminating the stop words from the corpus. Since they are correlated, this might be helpful in using only one of them, either the first letters or the last letter. Some letters have zero or less frequencies in both first and last letters (e.g. *b*, *j*, *q*, *v* and *t*). The highest frequency obtained for the last letters is for the letter *s* (more than *1000*) and the highest for the first letter is for the letter *e* (more than *600*).

To recap, the analysis provided in this research is very informative and clearly shows some patterns in the collection. These patterns can be useful for comparing different feature-outcome relationships. Finally, with clean data in place and an adequate comprehension of the content of the corpus resulted from the *EDA*, it is now a suitable time to test some modeling techniques on the data since the results give an indication of statistical significance about the data.

V. Modeling of the Collection

After data cleansing and good comprehension of the content as a result of the analysis, the data is now ready for:

- Building models with the goal of making better predictions concerning all parts of the SSRs related to PEOs.
- Implementing association rules on the PEOs.
- Gaining an understanding of the PEOs to make better decisions on all issues related to the PEOS.
- Classifying and clustering the PEOs.

This phase is much more focused than the EDA process, because the intended outcome is specific and clear. The classification method has been selected because it is one of the significant techniques in supervised learning [2]-[7]. In addition, the dataset is based on the classification and the analysis of the data support the classification. The classification of the PEOs is also one of the most critical and daunting tasks in building and evaluation of the SSRs. It is very important to divide the data into training and testing sets before implementing classification on the data set. In addition, the general nature of the data is classification (PEOs are mapped to SOs) and the analysis has clearly shown many patterns on the PEOs.

A. Training and Testing Sets

Table I illustrates some of the most important statistics and specification of the collection.

TABLE I
Summary of the Important Statistics of the Collection

No. of Insts.	No. of Attrs.	No. of Labels	Label Cardin.	Validation Method	No. Train. Insts.	No. Train. Insts.
167	160	11	3.8	Training/Testing split (66%, 34%)	110	57

The collection has been preprocessed to convert the data instances into a representation suitable for machine learning algorithms. Table II presents the specifications of the main preprocessing steps of the collection.

TABLE II
Specification of the Main Processing Steps Implemented on the Collection

Step	Specification
Tokenization	Alphabetic Tokenize
Lowercase Tokens	True
Stop words Handler	Rainbow
Stemmer	Snowball Stemmer
Min. Term Frequency	3
TF transform	True
IDF Transform	True
Normalization	All data

B. Measuring the Effectiveness of Techniques

In experiments, we used various evaluation measures that have been suggested by Tsoumakasetal [22]. More specifically, to evaluate different MLC methods, the following evaluation measures were used: Accuracy (Acc.), Hamming Loss (HL), Micro-F, Macro-F (Example-based), Macro-F (Label-based).

1. Machine Learning Techniques

All multi-label classification methods and supervised learning algorithms used in this work are implementations of the Weka-based [20] package of Java classes for multi-label classification, called Meka [23]. This package includes implementations of some of the multi-label classification methods most widely applied in the literature. All the algorithms were supplied with Weka's J48 implementation of a C4.5 tree classifier as a single-label base learner.

In the experiment, different multi-label classification methods have been used, where six are problem transformation methods (Binary Relevance (BR), Label Powerset (LP), Classifier Chains (CC), Pruned Sets (PS), and three are ensemble methods (Random k-label sets (RAkEL), Ensemble of Classifier Chains (ECC) and Ensemble of Pruned Sets (EPS)) and the remaining two are algorithms adaptation methods (Multi-Label k-Nearest Neighbors (ML-kNN) and Back-Propagation Multi-Label Learning (BPMLL)).

2. Parameter Configuration

All configurable parameters of the participating algorithms were set to their optimal values as reported in the relevant papers. For BR, LP and CC no parameters were required. The PS methods required two parameters p and strategy parameter for each dataset. We used p = 1 and strategy parameters, A_b, for all datasets with value of b = 2, as proposed by [24]. The number of models in the ECC methods was set to 10, as proposed by [25]. For RAkEL the number of models was set to 10 and the size of the label-sets K was set to half the number of labels. For EPS, at each dataset p and strategy, parameters were set to the same values as those used for the PS method. EPS requires additional parameter the number of models was set to 10. For all ensemble methods the majority voting threshold was set to 0.5.

VI. Results

Table III summarizes the results obtained using the above-mentioned evaluation measures.

The results of the experimental comparison revealed that the EPS and ECC perform better than the remaining methods in terms of accuracy, Micro-F1, Macro-F1 (By Example), and Macro-F1 (By Label). It is also obvious that the CC perform the lowest. Moreover, Table IV shows the results in terms of the accuracy per label.

TABLE III
PERFORMANCE OF MLC METHODS

	Acc.	**HL**	**Micro-F1**	**Macro-F1 (By Example)**	**Macro-F_1 (By Label)**
BR	0.274	0.391	0.442	0.389	0.427
LP	0.259	0.34	0.42	0.352	0.411
CC	0.264	0.325	0.37	0.335	0.363
PS	0.284	0.324	0.425	0.366	0.41
RAkEL	0.338	0.365	0.501	0.462	0.469
ECC	0.358	0.357	0.547	0.48	**0.534**
EPS	0.368	0.397	0.518	**0.49**	0.515
BPNN	0.281	0.37	0.37	0.377	0.456
MLkNN	0.32	0.351	0.5	0.45	0.481

TABLE IV
PER LABEL ACCURACY OF MLC METHODS

MLC method \ Label	**SO-a**	**SO-b**	**SO-c**	**SO-d**	**SO-e**	**SO-f**	**SO-g**	**SO-h**	**SO-i**	**SO-j**	**SO-k**
BR	0.754	0.632	0.684	0.351	0.667	0.702	0.667	0.719	0.667	0.298	0.561
LP	0.719	0.561	0.719	0.719	0.667	0.596	0.667	0.667	0.719	0.614	0.614
CC	0.719	0.702	0.684	0.754	0.702	0.684	0.632	0.614	0.684	0.649	0.596
PS	0.702	0.596	0.684	0.754	0.684	0.737	0.579	0.719	0.632	0.737	0.614
RAkEL	0.719	0.719	0.579	0.649	0.667	0.544	0.579	0.684	0.702	0.579	0.561
ECC	0.737	0.632	0.614	0.667	0.667	0.596	0.596	0.684	0.614	0.632	0.632
EPS	0.614	0.667	0.491	0.561	0.544	0.544	0.544	0.772	0.649	0.632	0.614
BPNN	0.719	0.579	0.526	0.684	0.632	0.509	0.649	0.719	0.667	0.684	0.561
MLkNN	0.719	0.702	0.632	0.719	0.649	0.544	0.579	0.702	0.649	0.649	0.596

Due to the small size of the collection, the above mentioned results cannot be considered as conclusive, but they give an initial impression about the multi-label classification in the context of educational data mining. The data collection can be extended and generalized by collecting more data and performing further investigation.

VII. Conclusion and Future Work

The available data collections are the driving force which are propelling the research in machine learning in general and supervised text classification in specific. The successful benchmark conducted on the collection presented in this research has the potential to substantially contribute in the advancement of the research in the text classification in accreditation processes and to shed light on different aspects of accreditation processes as a new research area related to text mining and machine learning.

The work presented in this research may also encourage other researchers to produce future data collections and managing real-world text classification systems such as collections related to mapping PEOs to program missions and visions. At the end, we hope that the benchmark collection we presented in this research will play a vital role in the automation of SSR evaluation.

References

[1] Fabrizio Sebastiani, Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47, 2002.
[2] Shweta C. Dharmadhikari, Maya Ingle, Parag Kulkarni, Empirical Studies on Machine Learning Based Text Classification Algorithms, Advanced Computing: An International Journal (ACIJ), Vol.2, No.6, November 2011.
[3] Daniel Carlos Guimarães Pedronette, Combined unsupervised and semi-supervised learning for data classification, 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Salerno, Italy, 13-16 Sept. 2016.
[4] Lunke Fei, Yong Xu , Xiaozhao Fang, and Jian Yang, Low rank representation with adaptive distance penalty for semi-supervised subspace classification, Pattern Recognition, Volume 67, Pages 252–262, July 2017.
[5] Bishop Christopher M., Pattern Recognition and Machine Learning, Springer, 2007.
[6] Murphy Kevin P., Machine Learning: A Probabilistic Perspective, The MIT Press, 2012.
[7] Duda Richar O., Hart PE, Stork DG, Pattern Classification. Wiley-Interscience, 2000.
[8] D. D. Lewis, R. E. Schapire, J. P. Callan, and R. Papka, Training algorithms for linear text classifiers, In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 96)*, pages 298–306, 1996.
[9] D. D. Lewis, Reuters-21578 text Categorization test collection. Distribution 1.0. README file (version 1.2). Manuscript, September 26, 1997.
[10] Y. Yang, An evaluation of statistical approaches to text categorization. *Information Retrieval*, 1 (1/2):67–88, 1999.
[11] David D. Lewis, Yiming Yang, Tony G. Rose, Fan Li RCV1: A New Benchmark Collection for Text Categorization Research, Journal of

Machine Learning Research 5: 361-397, 2004.
[12] P. Y. Pawar and S. H. Gawande, A Comparative Study on Different Types of Approaches to Text Categorization, *International Journal of Machine Learning and Computing*, 2-4:423-426, 2012.
[13] ABET, ABET Strategic Plan, Accreditation Board for Engineering and Technology, Inc., (ABET), November 1, 1997
[14] Engineering Accreditation Commission (ABET), Criteria for Accrediting Engineering Programs Effective for Review During the 2015-2016 Accreditation Cycle, 415 N. Charles Street Baltimore, MD 21201, United States of Ameriaca, ABET, 2014.
[15] ABET, Criteria for Accrediting Engineering Programs Effective for Reviews During the 2016-2017 Accrediting Cycle.
[16] Ronald Dekker, The Importance of Having Data-sets, Proceedings of the IATUL Conferences, Paper 16, May 23rd, 2006.
[17] Anwar Ali Yahya, Zakaria Toukal, Addin Osman, Bloom's Taxonomy–Based Classification for Item Bank Questions Using Support Vector Machines, Modern Advances in Intelligent Systems and Tools, SCI 431, pp 135-140, 2012.
[18] Anwar Ali Yahya and Addin Osman, Automatic Classification of Questions into Bloom's Cognitive Levels using Support Vector Machines, The International Arab Conference on Information Technology, At Naif Arab University for Security Science (NAUSS), Riyadh, Saudi Arabia, 2011.
[19] Anwar Ali Yahya, Addin Osman, Ahmed Abdu Alattab, Educational data mining: A case study of teacher's classroom questions, 13th International Conference on Intelligent Systems Design and Applications (ISDA), UPM, Selangor, Malaysia, 2013.
[20] D. Koller and M. Sahami. Hierarchically classifying documents using very few words. In International Conference on Machine Learning (ICML'97), pages 170–178, Nashville, 1997.
[21] A. S. Weigend, E. D. Wiener, and J. O. Pedersen. Exploiting hierarchy in text categorization. Information Retrieval, 1(3):193–216, 1999.
[22] G. Tsoumakas, I. Katakis, I. Vlahavas, Mining multi-label data, in: Data Mining and Knowledge Discovery Handbook, Springer, Berlin/Heidelberg, 2010, pp. 667–685.
[23] Jesse Read, Peter Reutemann, Bernhard Pfahringer, and Geoff Holmes. MEKA: A multi-label/multi-target extension to Weka. Journal of Machine Learning Research, 17(21):1–5, 2016. URL http://jmlr.org/papers/v17/12-164.html.
[24] I. H. Witten, ans E. Frank, ―Data Mining: Practical Machine Learning tools and techniques, Morgan Kaufmann, 2005.
[25] J. Read, B. Pfahringer and G. Holmes, ―Multi-label classification using ensembles of pruned sets‖, Proc 8th IEEE International Conference on Data Mining, Pisa, Italy, pages 995-1000. IEEE Computer Society, 2008.
[26] N. Abbadeni, A. Ghoneim, and A. Alghamdi, Program Educational Objectives Definition and Assessment for Quality and Accreditation, International Journal of Engineering Pedagogy, 3 (3):33-46, 2013.
[27] J. W. Tukey, Exploratory Data Analysis 1st Edition, Addison-Wesley, 1977.
[28] D. Cielen, A. D. B. Meysman, M. Ali, Using Python Tools, Introducing Data Science, Big Data, Machine Learning, and More, Manning Publication Co., Shelter Island, NY 11964, 2016.
[29] J. M. Stanton, J. M., Introduction to Data Science, Third Edition. iTunes Open Source eBook. Available: https://itunes.apple.com/us/book/introduction-to-data-science/id529088127?mt=11, 2012.
[30] S. Bird, E. Klein and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.