Data Mining in Business Education: Exploratory Analysis of Course Data and Job Market Requirements

Zinovy Radovilsky*
California State University, East Bay, Hayward, California, USA

Vishwanath Hegde
California State University, East Bay, Hayward, California, USA

Leena H. Damle
California State University, East Bay, Hayward, California, USA

The Data Mining (DM) subject is increasingly incorporated into business programs as a response to the growing demand for data analytics skills in the market. The evolving academic response to DM teaching has been wide-ranging because of the interdisciplinary nature of the DM subject. In this research, we reviewed and systemize the contents and skills relevant to the DM curricula. We also analyzed and contrasted DM courses that are offered at different layers of business education, evaluated the job market requirements and compared the two. The research results will enable more effective DM course teaching that is aligned with the job market requirements.

* Corresponding Author. E-mail address: zinovy.radovilsky@csueastbay.edu

I. INTRODUCTION

For a number of years data analytics has been showing a tremendous growth and proliferation in multiple fields and organizations. According to the Dresner Advisory Service' 2018 Big Data Analytics Market Study, a combined big data deployment in enterprises in various industries soared from 17% in 2015 to 59% in 2018 (Columbus, 2018). Increased interest from businesses, governments, and other organizations in harnessing the power of big data analytics and technology is corresponding with the need for well-trained professionals, to handle the related data analytics tasks. As a result, a number of analytics professions emerged, among which business data analytics (BDA) becomes one of the most prominent fields. The need of BDA specialists led to a substantial increase in the number of academic programs in analytics, and, in particular, master programs in business analytics. The number of graduate business analytics programs has grown from approximately 30 in 2014 to 110 programs in mid-2019, or more than 3.5 times (Rappa, 2019).

The Data Mining (DM) subject is increasingly incorporated into business programs as a response to the growing demand for analytics skills, either as a part of specialized business analytics programs or general business programs. According to a web mining study of AACSB-accredited U.S. colleges of business (Zhao and Zhao, 2016), the DM courses are one of the most frequent courses offered by the specialized MS and BA/BS programs in business analytics (#2 and #3 most frequent, respectively), and #1 most frequent course in MBA concentrations in business analytics. Data mining can be taught to students with different backgrounds and levels of knowledge (Gao and Zhang, 2018).

The evolving academic response has been quite varied because of the interdisciplinary nature of the DM subject and the developing data analytics landscape. This poses a number of challenges/issues in designing and teaching DM courses in different business programs and majors. The first major issue pertains to the identification of relevant contents and associated skills that should be incorporated into the DM courses. The second major issue is the emphasis attached to traditional analytics contents versus skills in DM courses. Finally, another major issue in developing DM courses is a gap existing between the contents and skills incorporated into the DM curricula and the market requirements for business analytics jobs.

The exploratory analysis in this paper was motivated by the need to better understand the contents and skills of DM courses from a variety of viewpoints, which would enable a more effective DM teaching amid the major issues stated above. To accomplish this research, we perform the following steps. First, based on literature review, we developed a set of contents and skills relevant to the DM curricula. Second, using these contents and skills, we compared and contrasted the DM courses that are offered at various layers of business education. Third, we also analyzed the job market requirements to identify their DM-related contents and skills. Finally, we compared the DM courses and job market requirements in terms of similarities and differences of their respective DM contents and skills. The main results of this research are associated with the identified contents and skills required in DM teaching, and the findings of the comparative analysis of the DM courses and the job market requirements. These will enable more effective DM course teaching that is aligned with the job market requirements.

The structure of this paper is as follows. After the Introduction, we present in section II a literature review of existing research on the DM curricula and their contents and skills. In section III, we describe the research methodology, summarize content- and skill-based categories of the DM curricula, present data collection and analytical approaches applied in this research. In section IV, we compare syllabi of the DM courses that are offered at different layers of business education. In section V, we analyze data mining requirements existing in modern BDA job market, and contrast the job market requirements with the results of DM course syllabi analysis. We finish the paper with the conclusions in section VI.

II. LITERATURE REVIEW

Data mining is a relatively new and emerging trend in educating business professionals, and hence the existing research in the area of DM curriculum is also evolving. In this section, we consider and analyze literature sources on developing DM courses and, specifically, their analytical content and skills generated through teaching those courses.

We identify a number of research papers that describe DM courses and their respective topics/content (Chakrabarti et al., 2006; Dokania and Kaur, 2018; Li, 2011; Luan, 2002; Wu et al., 2015). Typically, a core of *content-based categories* of a DM course that are presented in each of these papers include: (a) data processing, cleaning, and preparation; (b) regression (multiple linear regression and logistic regression); (c) classification and regression trees; (d) neural networks; (e) clustering; and (f) association analysis. However, each of these papers varies in terms of adding some other content-based categories related to a DM course. For example, Dokania and Kaur (2018) add to the specified core content-based categories several more categories such as time series analytics and k-nearest neighbors classification method. Li (2011) expands the core content-based categories with data warehousing and online analytical processing (OLAP), text and web mining, visual DM, industry efforts and social impacts. A detailed template of a DM curriculum described by Chakrabarti et al. (2006) enlarges the specified core content-based categories of a DM course with a number of methods and techniques like naïve Bayesian classification, frequent pattern analysis, linear and nonlinear classification, ensemble classification, outlier analysis, and some others.

Reviewing these research papers and reports, we recognize that, despite the suggestion that "the course content for data mining has been well defined and streamlined because of the availability of outstanding data mining textbooks" (Li, 2011), the actual content of DM courses may significantly fluctuate. Moreover, the analyzed papers, while describing content-based categories of a DM curriculum, have minimum to no consideration for all kinds of problem solving, decision making, managerial and communication skills that may be generated through teaching a DM course. We will refer to them as DM *skill-based categories*.

To further analyze content and skills generated through DM teaching, we reviewed design and development methodologies for a DM curriculum (Asamoah et al., 2017; Bowers et al., 2018; Jafar, 2010; Pan et al., 2018; Sanati-Mehrizy et al., 2015; Rawat, 2017; Tosic and Beeston, 2018). Jafar (2010) states that a DM course in an Information Systems program (a part of a business school) should be designed with three main components: (1) Analytical component, (2) Tool-based, hands-on component, and (3) Rich collection of data sets. The first analytical component covers the theory and practice of a DM analysis project, elementary data analysis, market basket analysis, classification and prediction, cluster analysis and category detection, testing and validation of mining models, and finally the application of mining models for decision support and prediction. The second hands-on component requires the use of tools to build projects based on the algorithms learned in the analytical component. The third component contains rich collection of data sets to demonstrate the capabilities of the algorithms and to provide practical hands-on experience, homework assignments, etc. (Jafar, 2010).

According to Rawat (2017), a computer science/engineering course in data mining consists of teaching five main applications: classification, regression, clustering, association rules, and anomaly/outlier detection. Such course may also incorporate a variety of data sets like relational databases, multimedia databases, special databases, time series databases, and some others (Rawat, 2017). A research by Sanati-Mehrizy et al. (2010) identifies, based on analyzing various *DM* courses from computer science departments, 4 main approaches in teaching data mining including:

(a) Mathematical/algorithm approach which involves the mathematics of how various algorithms are derived and applied; (b) Textbook approach that may have similar qualities to the mathematical/ algorithm based courses but is noticeably different in that they are structured around a specific text and follow it almost exclusively; (c) Topic approach, which does not follow a specific textbook nor is it primarily focused on the derivation and intricacies of the various algorithms used in DM; and (d) Applied DM approach that combines theory with hands-on application, i.e., it applies the concepts being taught to current real-world problems. Moreover, according to the specified survey, the first mathematical/algorithm approach is the most popular among the DM course in computer science departments (Sanati-Mehrizy et al., 2010).

Based upon the analysis of the literature sources thus far, we made several observations. First, the reviewed methodologies of designing and teaching DM are mostly concerned with presenting to students a range of DM content-based categories associated with specific DM methods and tools. Second, the designing and teaching of DM courses are approached differently that may potentially depend on a school/department offering this course, e.g., a course from management information systems/information technology department versus a course from business-oriented department. In addition, the differences between the methodologies can be due to a level of teaching such as an undergraduate DM course versus its graduate-based equivalent. Third, in the reviewed papers, we also did not find the discussion of skills that need to be developed through teaching DM. Finally, there was no assessment of the DM courses through the prism of existing business data analytics requirements in the job market.

We did not find from literature search any paper that would specifically address skill-based categories in DM courses. We identified, however, few papers that consider development of skill-based categories in data analytics programs and courses that may potentially refer to DM courses as well (Bowers et al., 2018; Phelps and Szabat, 2017; Radovilsky et al., 2018; Rienzo and Chen, 2018).

The most extensive research in this area thus far is done by Bowers et al. (2018). The authors utilize multiple job postings in analytics, big data, business analytics, and data science to identify most frequent terms (topics) in business data analytics and data scientist job postings. Among the most frequent job skills required in business analytics were (in descending order of the frequency percentage): communication and interpersonal skills (61.77%), business domain (36.90%), managerial skills (36.63%), database (26.03%), analytics tools (19.76%), system analysis and design (15.82%), modeling and analysis (8.88%) and some others of lesser frequencies. Bowers et al. (2018) also conducted a survey of graduate academic programs in business data analytics and data science in terms of the content and skills generated by these programs, and then compared the survey results with the results of topic frequencies from job postings. They found that communications and interpersonal skills and managerial skills are most frequent in the job posting skills, but, on the average, less than 5% of the academic programs are dedicated to so-called soft skills regardless of program type. Another major gap exists between the frequent job requirements of business domain and its limited coverage in existing business analytics programs (Browers et al., 2018).

Some authors just analyzed job market postings in analytics, including business data analytics (BDA) and data science (DS) jobs, to identify specific skills sets required for these fields. Rienzo and Chen (2018) examined 400 analytical job postings on Indeed.com. Their findings indicate that the top six skills for analytical jobs are SQL, Excel, project management, SAS, statistics, and database. They also recognize that communication was not included in job postings analysis, and that it, like project management, is a critical skill for analysts. However, this research stopped short of comparing these skills with the skills generated through teaching data analytics in academic courses. Another group of authors (Radovilsky et al., 2018) examined various online job sites to identify BDA and DS real job skills required from job markets. These skills were grouped into four knowledge domains: Technical, Analytical, Business, and Communication. Utilizing text mining methods, Radovilsky et al. (2018) identified some similarities as well as important differences between knowledge domains the groups of skills for BDA and DS. This research also states the importance of its results to teaching data analytics in academia, but does not provide a comparison with the existing skills in academic BDA and DS curriculum.

Overall, the analysis of the described literature sources in terms of DM curriculum indicate that they predominately concentrate on describing existing and newly designed content-based categories without proper consideration for DM skill-based categories. The literature review also revealed that teaching DM is not, in most cases, connected with the existing job market requirements in business and data analytics. In addition, we were not able to identify any research or practitioner paper that would analyze and compare content- and skill-based categories in teaching DM versus job market requirements in business data analytics.

III. METHODOLOGY

Based on our literature review in the previous section and taking into consideration the importance of DM curriculum in business data analytics (Zhao and Zhao, 2016), we consider the following questions to be answered in this research:

1. What are content-based and skill-based categories in teaching an academic DM curriculum in business data analytics?
2. What are frequencies of various categories taught in DM and how do they compare for various layers of business education, i.e., graduate vs. undergraduate programs, Business DM courses versus Management Information Systems (MIS) / InformationTechnology Management (ITM) DM courses, and DM courses in general business versus specialized business education?
3. What are groups of DM content and skills in data analytics' job market requirements?
4. How do these groups of content and skills compare with the content- and skill-based categories in the DM curricula? What are the commonalities and gaps between the two?

For addressing the specified questions, we develop a research model, visual representation of which is shown in Figure 1. This model contains the main steps of our research process, which are directly associated with the specified research questions.

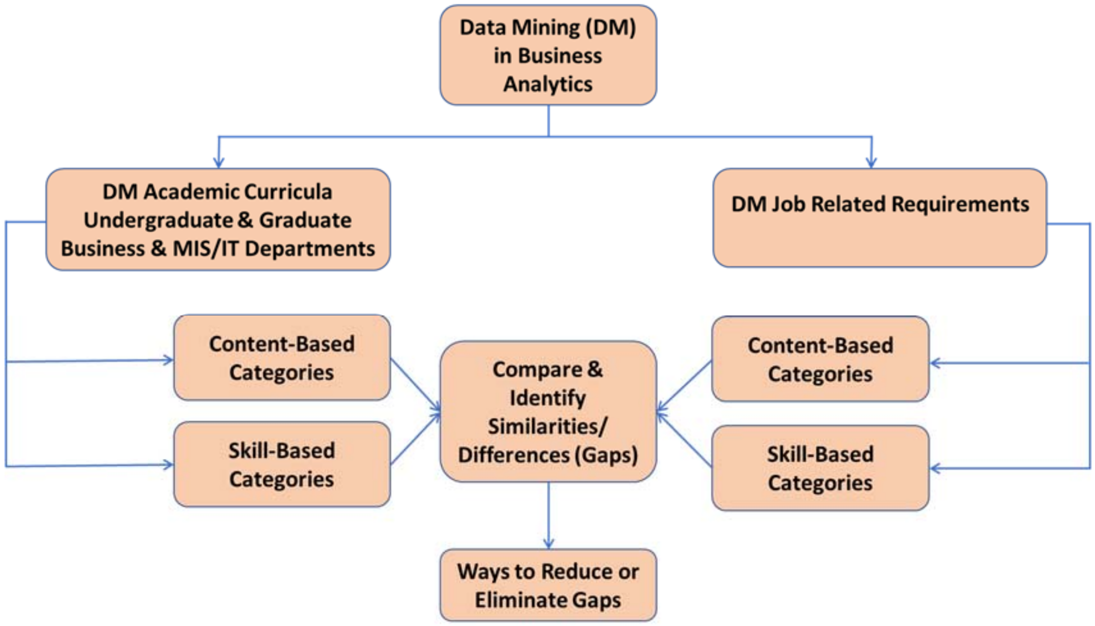

FIGURE 1. DATA MINING RESEARCH MODEL.

TABLE 1. DATA MINING CATEGORIES.

##	*Content-Based Categories*	*Skill-Based Categories*
1	Dataset Processing, Cleaning and Preparation	Using DM Software
2	Database Management and Data Warehousing	Programming Languages and Coding
3	Data Visualization	Problem Solving
4	Data Reduction/Principle Component Analysis	Apply DM Methods for Business Data
5	Multiple Regression	Selection of Best DM Methods
6	Logistic Regression	Modeling/Decision Making
7	Discriminant Analysis	Interpret and Present Results
8	k-Nearest Neighbors	Written and Oral Communication
9	Decision Trees	
10	Naïve Bayes/Bayes Models	
11	Support Vector Machine	
12	Neural Network	
13	Deep Learning	
14	Cluster Analysis	
15	Association/Affinity Analysis	
16	Collaborative Filtering/Recommender System	
17	Web Data Mining	
18	DM Ethics	
19	Text Mining	
20	Time Series Forecasting	
21	Ensemble Methods	
22	Outlier/Anomaly Detection	

To approach the first research question on DM contents and skills, we utilized information from 10 popular textbooks in data mining general list of DM content-based and skill-based categories (Aggarwal, 2010; Dean, 2014; Han et al., 2012; Larose and Larose, 2015; Ledolter, 2013; Provost and Fawcett, 2013; Shmueli et al., 2018; Tan et al, 2019; Torgo, 2011; Witten et al., 2011). We also use a variety of research papers from our literature review in the previous section to validate these categories (Asamoah et al., 2017; Bowers et al., 2018; Pan et al., 2018; Phelps and Szabat, 2017; Radovilsky et al., 2018; Rawat, 2017; Tosic and Beeston, 2018; Rienzo and Chen, 2018). Combining and summarizing information from textbooks and literature sources, we introduced 22 content-based categories and 8 skill-based categories of DM curriculum (Table 1). As researchers, we also reached consensus on these sets of content-based and skill-based categories in the table.

For answering the second research question, we collected and analyzed syllabi of various DM courses in business analytics and general business programs. All these syllabi were used to identify, using statistical analysis, the relative frequencies of content-based and skill-based categories (see Table 1), and compare them for different level of teaching: graduate vs. undergraduate programs, Business vs. Management Information Systems (MIS)/Information Technology Management (ITM) departments; and general vs. specialized education (see next section IV with the detailed analytical results).

To respond to the third research question on DM content and skills in the analytics job market, we analyzed BDA jobs that can be performed by those learning DM as a part of their Business Analytics programs. We developed a data set of 1,500 unique BDA job descriptions spanning over a period of about 3 years, from 2017 through 2019. A specific number of 1,500 records represents a substantial and sufficient array of records to be able to perform a qualified text mining analysis and receive statistically significant results of this analysis. The titles of these positions were somewhat different, for example, “Analyst,” “Data Analyst,” “Business Analyst,” “Data and Analytics Associate,” “Data Scientist,” ”Data Mining Analyst,” “Senior Business Analyst,” and “Lead Enterprise Analyst.” Despite different titles, we check job requirements for each position to confirm that it is directly related to business data analytics. The collection of these records was done manually without utilizing any web-scaping (web-crawling) tools, because with the latter tools we would not be able to structure the scrapped data into specific data/column structure we used in our research. They were derived from various online job posting sites including Indeed.com, LinkedIn.com, Monster.com, Dice.com, glassdoor.com, and careerbuilder.com.

For each job record in the specified data set, we collected the following information:

- Job Title – title of a job position by which it is advertised in the website.
- Education – education level required for the job.
- Experience – number of years of professional experience required for the job.
- Responsibilities – various business and workplace responsibilities and skills to be required from a prospective employee, e.g., teamwork, oral communication and writing capabilities, problem solving, etc.
- Requirements/Qualification/Skill s – The technical skills required

for that job role. This includes, for example, knowledge of certain software like Tableau and Excel, expertise in programming languages like Python and R, and techniques like data visualization, data mining/machine learning methods and so on.

In this research, we mostly concentrate on the data set columns with the information on job responsibilities, requirements, skills, and qualification. Assuming a huge array of words from 1,500 records that these columns contained, we utilize several text mining methods to identify the frequencies of DM content and skills in the BDA job requirements. For this, we utilize Document Term Matrix (DTM), term cloud, and term frequencies that are commonly used for text mining analysis and are well described in DM literature and textbooks (Kwartler, 2017; Shmueli et al., 2019; Weiss et al., 2015). We also utilize the R software for the specified text mining analysis. R is one of the common statistical packages used in business data analytics. Moreover, it contains several packages with text mining functions that can be used to run the described text mining methods for analyzing job requirements (see section V with the detailed results).

Finally, to address the fourth research question, we compare two sets of results: the first one derived from the groups of content and skills of the job requirements, and the second one – content-and skill-based categories identified from analyzing DM syllabi. Our ultimate objective in this comparison is to identify potential gaps that have to be alleviated to bring the DM curricula to the real job market requirements in business data analytics (see section V with the detailed discussion).

IV. ANALYSIS OF DATA MINING COURSE SYLLABI

Using an Internet search, we identified 89 syllabi of DM courses taught in specialized business analytics programs. All course syllabi came from the programs offered by 4-year academic universities in the U.S. and Canada that were identified from the existing lists of these programs (Discover Data Science, 2019; Study Portals, 2019). The considered syllabi have been developed for courses taught in a relatively short period of time from 2016 through 2019. Of the total number of syllabi, 58 of them (65% of the total) belong to the masters' programs in business analytics, and 31 (35% of the total) – to the specialized undergraduate programs in the same field. Of these 89 syllabi, 48 syllabi (54%) are offered by business departments, e.g., management, decision science, marketing, or finance; and 41 syllabi (46%) are offered by the management information systems (MIS) and information technology management (ITM) departments or groups of faculty.

With respect to the general business programs (MBA and undergraduate programs), we were able to identify a relatively small number of concentrations in business analytics (32 total) that introduce DM courses in their respective business programs, i.e., 14 course syllabi for undergraduate and 18 for graduate (MBA) concentrations in business analytics. Overall, the total number of syllabi that we considered in this research is equal to 121 (89 + 32) syllabi.

We analyzed each DM syllabi for content- and skill-based categories as presented in Table 1, and then tabulated these categories into a data set with rows representing the analyzed syllabi and columns describing DM content- and skill-based categories. Using this data set, we identified frequencies of specific content- and skill-based categories for the following types of course syllabi: (a) graduate and undergraduate DM course syllabi in business analytics programs; (b) course syllabi from business-related departments and MIS/ITM departments in business data analytics programs; and (c) DM course syllabi from specialized (graduate and undergraduate) business analytics programs and general business programs (graduate/MBA and undergraduate).

We start our analysis with the pairs of frequencies (in percentages) of content- and skill-based categories for graduate and undergraduate DM courses in business analytics programs (see Table 2). The data in Table 2 is sorted in descending order by the frequencies of the categories in the "Frequency in Graduate Courses, %" column. As can be seen from this data, the frequencies of some categories in both types of courses are very similar and positioned in the same order, for example, "Cluster Analysis," "Decision Trees," "Association/Affinity Analysis," "Multiple Regression" and some others. Contrary to that, some categories like "Neural Network," "Web Data Mining" or "Modeling/Decision Making" have substantially different frequencies between course syllabi in the graduate and undergraduate DM courses.

For further analysis of similarities and differences in frequencies, we utilized two-tail hypothesis testing for the differences between population proportions (population frequencies) for each content- and skill-based category in Table 2. A standard approach for hypothesis testing of the proportions' differences is well explained in a variety of sources (see, for example, Evans, 2013). For each category, the null hypothesis is that the respective proportions (frequencies) for this category in graduate (P_1) and undergraduate (P_2) DM course syllabi are equal: H_o: $P_1 = P_2$, and alternative hypothesis: H_a: $P_1 \neq P_2$. To test the null hypothesis, we calculate the pooled sample proportion, compute the standard error of the sampling distribution difference between two proportions, and identify the Z-score test statistic and associated P-value. We also use a significance α value of 0.05.

The hypothesis testing results, i.e., the Z-score test statistics and P-values are presented in Table 2. Any P-value of greater than 0.05 indicates that we may not reject the null hypothesis; hence, the DM category population frequencies for the graduate and undergraduate syllabi are the same. As can be seen from Table 2, a substantial portion of these categories' frequencies (17 categories out of the total of 30 categories) are statistically the same, which means that, in many respects, the DM graduate and undergraduate syllabi are similar in terms of what contents and skills they introduce for teaching DM.

However, there is a number of categories in Table 2, for which the P-values are less than 0.05 (numbers are in bold in Table 2). This implies that we have to reject the null hypotheses, and, therefore, the respective population frequencies are different. For example, for such content categories as "Neural Network," "Text Mining," "Support Vector Machine," "Ensemble Methods" and "Web Data Mining," the frequencies of their presence in graduate syllabi is statistically larger than that for the undergraduate courses. This actually means that the graduate courses, based on their syllabi, teach more advanced DM topics as opposed to those in undergraduate courses. Contrary, the content categories like "Dataset Processing, Cleaning and Preparation," "Database Management and Data Warehousing," and "Time Series Forecasting" are, unconventionally, more frequent in the undergraduate courses as opposed to graduate ones. Considering the skill-based categories and results of the hypotheses testing of their frequencies, we can conclude that the graduate syllabi statistically provide more attention (more frequent) for "Modeling/Decision Making," "Interpret and Present Results," and "Written and Oral Communication" (P-values are less than 0.05; are in bold in Table 2).

Based on the original data set, we also analyzed the frequencies of content- and skill-based categories for DM syllabi coming from business departments (Management, Marketing, Finance, etc.) vs. information technology departments (MIS, ITM, etc.). The frequencies and hypothesis testing results (as previously explained in this section of the paper) are provided in Table 3.

TABLE 2. COMPARISON AND HYPOTHESIS TESTING OF DM CATEGORIES IN GRADUATE AND UNDERGRADUATE COURSE SYLLABI.

Categories	*Frequency in Graduate Courses, %*	*Frequency in Undergrad Courses, %*	*Test Stat. (Z-score)*	*P-value*
Content-Based Categories				
Cluster Analysis	94.8	96.8	-0.420	0.731
Decision Trees	87.9	83.9	0.531	0.693
Association/Affinity Analysis	75.9	77.4	-0.164	0.787
k-Nearest Neighbors	**65.5**	**32.3**	**2.981**	**0.009**
Naïve Bayes/Bayes Models	51.7	54.8	-0.279	0.767
Multiple Regression	51.7	67.7	-1.447	0.280
Neural Network	**48.3**	**19.4**	**2.659**	**0.023**
Text Mining	**46.6**	**19.4**	**2.515**	**0.034**
Support Vector Machine	**41.4**	**12.9**	**2.739**	**0.019**
Data Visualization	36.2	29.0	0.678	0.634
Logistic Regression	34.5	29.0	0.519	0.697
Dataset Processing, Cleaning and Preparation	**34.5**	**64.5**	**-2.697**	**0.021**
Ensemble Methods	**29.3**	**6.5**	**2.492**	**0.036**
Data Reduction/Principle Component Analysis	27.6	25.8	0.180	0.785
Collaborative Filtering/Recommender System	27.6	29.0	-0.144	0.790
Web Data Mining	**24.1**	**0.0**	**2.961**	**0.010**
Outlier/Anomaly Detection	**19.0**	**0.0**	**2.574**	**0.029**
DM Ethics	17.2	22.6	-0.607	0.664
Deep Learning	12.1	16.1	-0.531	0.693
Database Management and Data Warehousing	**10.3**	**32.3**	**-2.549**	**0.031**
Time Series Forecasting	**8.6**	**32.3**	**-2.821**	**0.015**
Discriminant Analysis	5.2	16.1	-1.711	0.185
Skill-Based Categories				
Modeling/Decision Making	**36.2**	**9.7**	**2.670**	**0.023**
Using DM Software	34.5	45.2	-0.982	0.493
Interpret and Present Results	**31.0**	**6.5**	**2.631**	**0.025**
Apply DM Methods for Business Data	25.9	22.6	0.340	0.753
Programming Languages and Coding	25.9	19.4	0.685	0.631
Problem Solving	20.7	16.1	0.518	0.698
Selection of Best DM Methods	19.0	16.1	0.330	0.756
Written and Oral Communication	**17.2**	**0.0**	**2.439**	**0.041**

TABLE 3. COMPARISON AND HYPOTHESIS TESTING OF DM CATEGORIES IN COURSE SYLLABI FROM BUSINESS AND MIS/ITM DEPARTMENTS.

Categories	*Frequency in Business Courses, %*	*Frequency in MIS/ITM Courses, %*	*Test Stat. (Z-score)*	*P-value*
Content-Based Categories				
Cluster Analysis	91.7	95.1	-0.648	0.647
Association/Affinity Analysis	**85.4**	**51.2**	**3.498**	**0.002**
Decision Trees	83.3	87.8	-0.595	0.668
Multiple Regression	60.4	51.2	0.872	0.546
k-Nearest Neighbors	56.3	43.9	1.161	0.407
Neural Network	52.1	58.5	-0.610	0.662
Logistic Regression	**52.1**	**24.4**	**2.666**	**0.023**
Dataset Processing, Cleaning and Preparation	45.8	56.1	-0.965	0.501
Text Mining	41.7	56.1	-1.358	0.317
Naïve Bayes/Bayes Models	33.3	43.9	-1.023	0.473
Collaborative Filtering/Recommender System	**33.3**	**9.8**	**2.656**	**0.023**
Time Series Forecasting	**29.2**	**7.3**	**2.614**	**0.026**
Data Visualization	25.0	36.6	-1.185	0.395
Support Vector Machine	**20.8**	**43.9**	**-2.336**	**0.052**
Database Management and Data Warehousing	20.8	29.3	-0.919	0.523
Data Reduction/Principle Component Analysis	18.8	29.3	-1.165	0.405
DM Ethics	18.8	22.0	-0.375	0.744
Discriminant Analysis	12.5	7.3	0.808	0.575
Ensemble Methods	**8.3**	**26.8**	**-2.323**	**0.054**
Web Data Mining	6.3	9.8	-0.612	0.661
Deep Learning	**0.0**	**22.0**	**-3.424**	**0.002**
Outlier/Anomaly Detection	**0.0**	**12.2**	**-2.490**	**0.036**
Skill-Based Categories				
Modeling/Decision Making	**50.0**	**17.1**	**3.250**	**0.004**
Apply DM Methods for Business Data	**41.7**	**4.9**	**4.010**	**0.000**
Problem Solving	**41.7**	**7.3**	**3.690**	**0.001**
Using DM Software	37.5	34.1	0.329	0.756
Interpret and Present Results	33.3	26.8	0.665	0.639
Selection of Best DM Methods	**31.3**	**9.8**	**2.467**	**0.038**
Written and Oral Communication	**25.0**	**0.0**	**3.442**	**0.002**
Programming Languages and Coding	**8.3**	**29.3**	***-2.564***	**0.030**

Comparable to data in Table 2, there is also an extensive number of content- and skill-based categories (16 categories out of 30 total) in DM course syllabi offered from business and MIS/ITM departments that have statistically the same population frequencies (P-value is greater than 0.05). However, the hypothesis testing also demonstrates that there are statistically significant differences in frequencies in 8 out of 22 content-based categories and 6 out 8 skill-based categories (numbers are in bold in Table 3). For example, the content-based categories of "Support Vector Machine," "Ensemble Methods," "Deep Learning" and "Outlier/Anomaly Detection" are substantially more frequent in

MIS/ITM syllabi than those categories in Business Department syllabi. Contrary, the content categories of "Association/Affinity Analysis," "Logistic Regression," "Collaborative Filtering/Recommender System," and "Time Series Forecasting" are more frequent in business DM syllabi vs. these categories frequencies in MIS/CS/IT departments. Most of the skill-based categories (5 out of 8) are surprisingly more frequent in business vs. information technology departments. Overall, this analysis demonstrates the DM syllabi coming from business departments are more focused on presenting business-driven content and skills of data analytics, whereas information technology departments' syllabi are more focused on teaching advanced machine learning methods.

Finally, we present in this section the comparison of DM categories in course syllabi that belong to specialized business programs like M.S. or B.S. in Business Analytics, and general business programs, e.g., MBA or B.S./B.A. in Business Administration (see Table 4). In this case, we do not make comparison categories between only graduate and undergraduate courses in each respective program, mostly due to the that fact of the relatively low number of course syllabi identified for general business programs (18 for graduate and 14 for undergraduate syllabi).

As can be seen from the data in Table 4, the majority of population frequencies (15 out of 22) of content-based categories for course syllabi in specialized and general business programs are statistically the same (P-value greater than 0.05). However, for 7 content-based categories, the population frequencies in the specialized course syllabi is statistically higher than that in the general business programs (numbers are in bold in Table 4). These categories include: "Text Mining," "Neural Network" (marginally accepted in this list with P-value of 0.051), "Support Vector Machine," "Data Reduction/Principle Component Analysis," "Ensemble Methods," "Web Data Mining," and "Deep Learning" (marginally accepted in this list with P-value of 0.058). This list of topics clearly indicates that course syllabi in specialized programs provide more focus on machine learning and advanced topics in business data analytics as opposed to DM syllabi for general business programs. For the skill-based categories, "Using DM Software" and "Programming Languages and Coding" are more frequent in specialized course syllabi; whereas "Apply DM Methods for Business Data," "Interpret and Present Results,“ and "Problem Solving" are unexpectedly more prevalent in general business courses.

TABLE 4. COMPARISON AND HYPOTHESIS TESTING OF DM CATEGORIES IN COURSE SYLLABI IN SPECIALIZED PROGRAMS AND GENERAL BUSINESS PROGRAMS.

Categories	*Frequency in Specialized Programs, %*	*Frequency in General Business Programs, %*	*Test Stat. (Z-score)*	*P-value*
Content-Based Categories				
Cluster Analysis	95.5	87.5	1.563	0.235
Decision Trees	86.5	84.4	0.299	0.763
Association/Affinity Analysis	75.3	81.3	-0.687	0.630
Text Mining	**57.4**	**28.1**	**2.844**	**0.014**
k-Nearest Neighbors	53.9	56.3	-0.226	0.778
Multiple Regression	53.9	43.8	0.988	0.490
Naïve Bayes/Bayes Models	52.8	40.6	1.182	0.397
Dataset Processing, Cleaning and Preparation	44.9	43.8	0.117	0.792
Neural Network	**38.2**	**15.6**	**2.344**	**0.051**
Data Visualization	33.7	40.6	-0.701	0.624
Support Vector Machine	**33.7**	**6.3**	**3.020**	**0.008**
Logistic Regression	32.6	43.8	-1.132	0.421
Collaborative Filtering / Recommender System	28.1	25.0	0.337	0.754
Data Reduction/Principle Component Analysis	**26.1**	**0.0**	**3.214**	**0.005**
Ensemble Methods	**21.3**	**0.0**	**2.847**	**0.014**
DM Ethics	19.1	28.1	-1.066	0.452
Time Series Forecasting	16.9	12.5	0.581	0.674
Database Management and Data Warehousing	15.7	6.3	1.358	0.317
Web Data Mining	**15.7**	**0.0**	**2.386**	**0.046**
Deep Learning	**14.6**	**0.0**	**2.288**	**0.058**
Outlier/Anomaly Detection	12.4	3.1	1.499	0.259
Discriminant Analysis	9.0	6.3	0.483	0.710
Skill-Based Categories				
Using DM Software	**39.1**	**15.6**	**2.427**	**0.042**
Modeling/Decision Making	26.1	25.0	0.121	0.792
Programming Languages and Coding	**23.6**	**0.0**	**3.023**	**0.008**
Apply DM Methods for Business Data	**17.4**	**40.6**	**-2.657**	**0.023**
Interpret and Present Results	**17.4**	**37.5**	**-2.329**	**0.053**
Problem Solving	**13.0**	**43.8**	**-3.647**	**0.001**
Selection of Best DM Methods	13.0	28.1	-1.946	0.120
Written and Oral Communication	13.0	15.6	-0.364	0.747

V. ANALYSIS OF DATA MINING CONTENT AND SKILLS IN JOB REQUIREMENTS

As discussed in section III on Methodology, we utilized text mining methods with *R's tm* package to analyze BDA job responsibilities, requirements, qualifications, and skills from the developed data set with 1,500 job records. The words extracted from the jobs were stored in a structured format referred to as *text corpus*, and we cleaned the corpus to isolate words not relevant to our research.

Based on the cleaned text corpus, we developed a *Document Text Matrix (DTM)*, which is a document containing actual count and relative frequency of terms in our job data set. The development of DTM allows to use, besides individual terms, the phrase terms combined of two or more words. The *Term Cloud* in Figure 2, produced by R's *tm* package, is a visual representation of the terms and their frequencies in the DBA job requirements, qualifications, and skills.

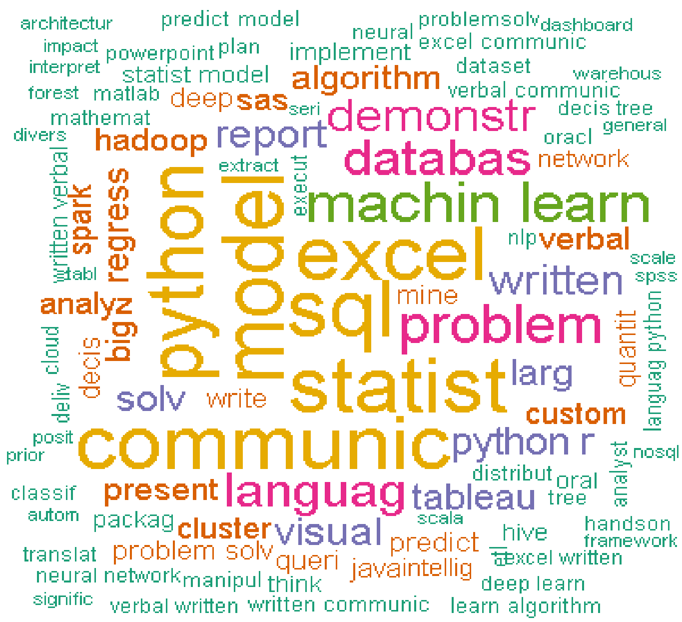

FIGURE 2. TERM CLOUD FOR BUSINESS DATA ANALYTICS JOBS.

As can be seen from the *Term Cloud*, the DM content- and skill-based categories form an important part of BDA job requirements. Among the most visible DM content and skill terms are software languages such as Python, R, SQL, and also software applications like Excel, Tableau, SAS, and some others. The Term Cloud also includes an extensive number of DM methods (content-based categories) like "Regression," "Clustering," "Neural Network," "Deep Learning," etc. The cloud also contains a substantive number of DM's skill-based categories like "Communication," "Reporting," "Demonstration Skills," "Problem Solving" and some other skills. In addition, the Term Cloud incudes some general terms directly associated with DM, e.g., "Machine Learning," Statistics," and "Predictive Models."

To further analyze the DM content-based categories in the job requirements, we identified frequencies of these terms using the "Requirements/Qualifications/Skills" corpus. The 20 most frequent DM content-based terms are presented in Tables 5.

The frequencies of job requirements' terms in the table reaffirm the importance of such DM content categories as "Database Management" (frequency of 38.5%), "Data Visualization" (27.4%), "Regression Analysis" (20.5%), "Cluster Analysis" (17.6%), "Deep Learning" (13.3%) and some others. At the same time, it may be hard to precisely identify frequencies of these and some other DM content categories, e.g., "Classification," "Text Mining," "Time Series Forecasting," "Association Analysis," etc. The main reason for this observation is a possibility that those DM categories may be incorporated into general BDA terms like

"Statistics" (73.1% of frequency), "Machine Learning" (43.8%), "Predictive Models" (22.5%), and "Data Mining" (16.9%).

To analyze skill-based categories in the job requirements, we present below Table 6 with 20 most frequent DM skill-based terms from the "Responsibilities" corpus.

TABLE 5. 20 MOST FREQUENT CONTENT-BASED TERMS.

##	*Term*	*Interpretation*	*Frequency, %*
1	statist	Statistics/Statistical Analysis	73.1
2	machin learn	Machine Learning	43.8
3	databas	Database/Database Management	38.5
4	visual	Visualization/Data Visualization	27.4
5	larg	Large Data Set	24.3
6	predict	Predictive Models/Analytics	22.5
7	regress	Regression/Regression Analysis	20.5
8	big	Big Data	20.0
9	execut	Execution/Model or Program Execution	19.7
10	cluster	Clustering/Cluster Analysis	17.6
11	mine	Mining/Data Mining	16.9
12	metric	Metrics for Model Performance	16.8
13	queri	Queries/Writing Queries	16.8
14	trend anal	Trends/Trend Analysis/Forecasting	16.6
15	evalu	Evaluation/Model Evaluation	15.9
16	deep	Deep Learning	13.3
17	quantit	Quantitative Models/Analysis	12.9
18	network	Neural Network	12.5
19	intellig	Intelligence/Artificial Intelligence	12.5
20	Decis tree	Decision Trees	10.3

TABLE 6. 20 MOST FREQUENT SKILL-BASED TERMS.

##	*Term*	*Interpretation*	*Frequency, %*
1	report	Report/Reporting	85.4
2	model	Models/Modeling	76.7
3	sql	SQL	75.7
4	communic	Communication Skills	71.8
5	excel	Excel	69.7
6	python	Python	51.6
7	languag	Languages/Programming Languages	33.5
8	demonstr	Demonstration Skills	30.4
9	written	Written Communication	30.1
10	problem	Problem Solving	28.4
11	python r	Python and R	26.9
12	tableau	Tableau	26.7
13	present	Presentation/Presentation Skills	25.1
14	implement	Implementation/Model Implementation	21.0
15	verbal	Verbal Communication	20.5
16	sas	SAS/SAS Applications	19.9
17	decis	Decision(s)/Decision Making	18.6
18	hadoop	Hadoop	18.3
19	spark	Spark	18.1
20	code write	Code Writing/Coding	17.1

The data in the table shows high frequencies of a number of BDA job terms related to communication skills, for example, "Reporting" (85.4% of frequency), "Written Communication" (30.1%), "Presentation" (25.1%) and "Verbal Communication" (20.5%). The table's data also demonstrates high frequencies for software skills, e.g., SQL (75.7%), Excel (69.7%), Python/Python and R (78.5% of combined frequency), and Tableau (26.7%). In addition, the table's frequencies display high proportion of terms related to "Models/Modeling" (76.7%), "Problem Solving" (28.4%), "Model Implementation" (21.0%), and "Decision Making" (18.6%).

The comparison of frequencies of various DM content- and skill-based categories from the course syllabi and BDA job requirements is presented in Table 7. The fist column represents the content- and skill-based categories that were common to both DM syllabi and job requirements' terms. Many other categories from DM syllabi that were analyzed in section IV of this paper, were not identified from the BDA job requirements, and thus could not be compared.

As can be viewed from this table, there are substantially higher frequencies of certain content-based categories derived from syllabi versus job requirements' terms, specifically, for "Cluster Analysis" (78.2% of difference), "Multiple Regression" (39.2%), and "Neural Network" (21.3%). Conversely, for "Database Management & Data Warehousing," the frequency from job requirements is higher than that from syllabi, -17.2% of difference.

TABLE 7. COMPARISON OF DATA MINING CATEGORIES IN SYLLABI AND JOB REQUIREMENTS

Data Mining Categories	*Average Frequency from Syllabi, %*	*Frequency from Job Requirements, %*	*Difference, %*
Content-Based Categories			
Cluster Analysis	95.8	17.6	78.2
Decision Trees	85.6	10.3	75.3
Multiple Regression	59.7	20.5	39.2
Neural Network	33.9	12.5	21.3
Data Visualization	32.6	27.4	5.2
Database Management & Data Warehousing	21.3	38.5	-17.2
Time Series Forecasting	20.5	16.6	3.9
Deep Learning	14.1	13.3	0.8
Skill-Based Categories			
Modeling/Decision Making	23.0	76.7	-53.8
Using DM Software	39.9	78.5*	-38.6
Programming Languages and Coding	22.7	50.7**	-28.0
Interpret and Present Results	18.8	25.1	-6.3
Problem Solving	18.4	28.4	-10.0
Written and Oral Communication	8.6	50.6	-42.0

* For category "Using DM Software," the frequency from job requirements is the total of frequencies from terms "Python" and "Python r" in Table 6.

** For category "Programming Languages and Coding," the frequency from job requirements includes a combination of frequencies from terms "language" and "code write" in Table 6.

Considering skill-based categories, the results in Table 7 clearly show that job requirements have noticeably higher frequencies for a number of categories versus respective frequencies of the same categories in DM syllabi. This is specifically relevant for "Modeling/Decision Making," "Using DM Software," "Programming Languages and Coding" and "Written and Oral Communication."

VI. CONCLUSIONS

The main objective of this paper was to analyze and compare the DM curricula content and skills existing at different layers of business education and also with the existing DM requirements in BDA job market in order to identify their similarities and differences. We summarize the paper findings into four main groups of results.

6.1. Recognize DM Literature Gaps and Introduce DM Content/Skill-based Categories

Overall, the analysis of the literature sources in terms of DM curriculum indicated that they predominately concentrate on describing existing and newly designed content-based categories without proper consideration for DM skill-based categories. The literature review also revealed that teaching DM is not, in most cases, connected with the existing BDA job market requirements. In addition, we were not able to identify any research or practitioner paper that would analyze and compare content- and skill-based categories in teaching DM vs. BDA job market requirements. The literature sources enabled us to introduce 22 content-based categories and 8 skill-based categories of DM curriculum that we apply for DM syllabi analysis.

6.2. Identify Differences of DM Curricula at Various Layers of Education

We analyze and compared, frequencies of content- and skill-based categories of DM syllabi at the three layers of business analytics education: (a) graduate and undergraduate DM course syllabi in business analytics programs; (b) course syllabi from business-related departments and MIS/ITM departments in BDA programs; and (c) DM course syllabi from specialized (graduate and undergraduate) business analytics programs and general business programs (graduate/MBA and undergraduate). Based on hypothesis testing, we identified that, although an extensive portion of these categories' frequencies is statistically the same, there is a substantive number of content- and skill-based categories of the DM curricula that are different at various layers of business analytics education.

Comparing graduate and undergraduate DM course syllabi in business analytics programs, we found that the graduate courses teach more frequently advanced DM content-based topics, e.g., neural networks, text mining, and ensemble methods. The graduate syllabi also placed more attention on some skill-based categories related to modeling and decision making, written and oral communication, and interpretation of DM results. Surprisingly, the content categories that involve data preparation, database management and data warehousing were more frequent in the undergraduate courses as opposed to graduate ones.

The comparison of DM course syllabi from business-related departments and MIS/ITM departments demonstrated that, in general, the syllabi coming from business departments are more focused on presenting business-driven content of data analytics, whereas MIS/ITM departments' syllabi are more focused on teaching advanced machine learning methods. At the same time, the business departments had unconventionally high frequencies of applying skill-based categories in their DM syllabi as opposed to these categories in MIS/ITM departments' syllabi.

Finally, the comparison of DM course syllabi from specialized business analytics and general business programs showed that the specialized programs' syllabi provide more focus on machine learning and advanced BDA topics in as opposed to DM syllabi for general business programs. For the most skill-based categories, however, the general business programs provide an unexpectedly high frequencies of various skills vs. those in specialized programs.

6.3. Systemize DM Contents and Skills from Job Market Requirements

The text mining analysis of a large number of BDA job records allowed us to identify DM content- and skill-based categories in the job market requirements. The frequencies of job requirements' terms reaffirmed the importance of several DM content categories such as database management, data visualization, regression analysis, cluster analysis, and deep learning. However, it was surprisingly hard to precisely identify frequencies of these and some other DM content-based categories, because they may be incorporated into general BDA terms like statistics, machine learning, predictive models, etc. Moreover, the BDA job requirements is commonly prepared by Human Resources professionals and job recruiters, and thus they may be applying more generic terms, instead of specific DM-associated categories. The DTM data for skill-based categories of the job requirements show high frequencies of a number of BDA job terms related to communication and reporting skills, software skills, modeling and problem solving.

6.4. Recognize Differences of Contents and Skills in DM Curricula and Job Market Requirements

We made some important conclusions by comparing frequencies of various DM content- and skill-based categories from the course syllabi and BDA job requirements. The DM syllabi demonstrated substantially higher frequencies of a number of content-based categories, e.g., cluster analysis, regression, and neural networks, as opposed to those in the job market requirements. Conversely for the skill-based categories, the results show that job market requirements provided noticeably higher frequencies for a number of skill categories vs. respective frequencies of the same categories in DM syllabi. The identified gaps can be an important point to consider while improving DM education in academia.

6.5. Research Contributions and Future Development

Overall, this research produced a number of important contributions to the theory and practice of DM teaching in business analytics and general business programs. First, we summarized, based on existing textbooks and literature sources, the DM curricula into a list of 22 content-based and 8 skill-based categories that are used in teaching DM. Second, we presented a unique research study, where content- and skill-based categories of the DM curricula are analyzed and compared at the three described layers of business analytics education. Third, we provide unique research results of identifying and systemizing the groups of DM content and skills used in BDA job market requirements. Finally, we compare the content- and skill-based categories of the DM curricula with those existing in the job requirements, identify their similarities and gaps.

This research can be extended in the future in several ways. We may utilize the research results for providing specific recommendations and solutions in designing new and improving existing DM courses for specific business analytics and general business programs. We may analyze the content- and skill-based DM categories in conjunction with several important job-related characteristics and attributes, i.e., years of experience, education, job location, and some others. In addition, we can analyze and compare DM teaching at various layers of business analytics education utilizing the proportion of time and depth of teaching specific content- and skill-based DM categories.

REFERENCES

Aggarwal, C. C., *Data Mining: The Textbook*, Springer, 2015.

Asamoah, D. A., Sharda, R., Zadeh, A. H., and Kalgorta, P., "Preparing a Data Scientist: A Pedagogic Experience in Designing a Big Data Analytics Course", *Decision Sciences Journal of Innovative Education*, 15(2), April 2017, 161-190.

Bowers, M. R., Camm, J. D., and Chakraborty, G., "The Evolution of Analytics and Implications for Industry and Academic Programs", *INFORMS Journal on Applied Analytics*, 48(6), November-December 2018, 487-499.

Chakrabarti, S., Ester, M., Fayyad, U., Gehrke, J., Han, J., Morishita, S., Piatetsky-Shapiro, G. and Wang, W., "Data Mining Curriculum: A Proposal (Version 1.0)", *The Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining*, May 2006, https://www.kdd.org (accessed June 10, 2019).

Columbus, L., "Where big data jobs will be in 2016", *Forbes*, Nov. 2015, http://www.forbes.com/sites/louiscolumbus/2015/11/16/where-big-data-jobs-will-be-in-2016/#358a4070f7f1 (accessed August 25, 2019).

Columbus, L., "Big Data Analytics Adoption Soared in the Enterprise in 2018", *Forbes*, https://www.forbes.com/sites/louiscolumbus/2018/12/23/big-data-analytics-adoption-soared-in-the-enterprise-in-2018/#57e9a728332f (accessed September 1, 2019).

Discover Data Science, "Master's in Business Analytics – Guide to Choosing a Program", *ttps://www.discoverdatascience.org/related-programs/masters-in-business-analytics/* (accessed August 02, 2019).

Dean, J., *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*, Wiley, 2014.

Dokania, N. K. and Kaur, N., "Comparative Study of Various Techniques in Data Mining", *International Journal of Engineering Sciences & Research Technology*, 7(5), May 2018, 202-209.

Evans, J. R. *Statistics, Data Analysis, and Decision Modeling*, Fifth Edition, Pearson, 2013.

Gao, Y. and Zhang, S., "Design of and Research on Autonomous Learning System for Distance Education based on Data Mining Technology", *Educational Sciences: Theory & Practice,* 18(6), 2017, 2633-2640.

Han, J., Kamber, M., and Jian, P., *Data Mining: Concepts and Techniques*, 3rd edition, Morgan Kaufmann, 2012.

Jafar, M. J., "A Tools-Based Approach to Teaching Data Mining Methods", *Journal of Information Technology*

Education: Innovations in Practice, 9(1), 2010, 1-24.

Kwartler, T., *Text Mining in Practice with R*, John Wiley & Sons Ltd, 2017.

Larose, D. T. and Larose, C. D. *Data Mining and Predictive Analytics*, 2nd Edition, Wiley, 2015.

Ledolter, J., *Data mining and Business Analytics with R*, Wiley, 2013.

Li, D., "A Practice-Oriented Approach to Teaching Undergraduate Data Mining Course", *American Society for Engineering Education*, 2011, 1-11.

Luan, J., "Data Mining and Its Applications in Higher Education", *New Directions for Institutional Research*, No. 113, Spring 2002, 17-36.

Pan, K., Blankley, A. I., Mazzei, Matthew J., Lohrke, Cynthia Frownfelter, Marshall, Jennings B. and Carson, Charles M., "Surveying industry advisors to select data analytics topics for all business majors", *The International Journal of Management Education, 16(3), November 2018, 483-492.*

Phelps, A. L. and Szabat, K. A., "The Current Landscape of Teaching Analytics to Business Students at Institutions of Higher Education: Who is Teaching What?", *The American Statistician*, 71(2), 2017, 155-161.

Provost, F. and Fawcett, T. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, O'Reilly Media, Inc., 2013.

Rappa M., "Graduate degree programs in analytics and data science", *https://analytics.ncsu.edu/?page_id=4184* (Accessed July 26, 2019).

Rawat, K. S., "Comparative Analysis of Data Mining Techniques, Tools and Machine Learning Algorithms for Effective Data Analytics", *IOSR Journal of Computer Engineering*, 19(4), July-August 2017, 56-61.

Rienzo, T. and Chen, K., "Planning for low end analytics disruptions in business school curricula", *Decision Sciences Journal of Innovative Education*, 16(1), 2018, 23–41.

Sanati-Mehrizy, R., Sanati-Mehrizy, P., Minaie, A. and Dean, C., "How a Data Mining Course Should Be Taught in an Undergraduate Computer Science Curriculum", *American Society for Engineering Education*, 2010, 1-7.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R. and Lichtendahl, K. C., *Data Mining for Business Analytics: Concepts, Technologies and Applications in R*, Wiley, 2018.

Study Portals, "Business Degrees in Business Intelligence & Analytics in United States", retrieved from https://www.bachelorsportal.com/study-options/269779178/business-intelligence-analytics-united-states.html, August 2, 2019.

Tan, P., Steinbach, M., Karpatne, A. and Kumar, V., *Introduction to Data Mining*, 2nd Edition, Pearson, 2019.

Tosic, P. T. and Beeston, J., "Designing Undergraduate Data Science Curricula: A Computer Science Perspective", *Proceedings of the ASEE Annual Conference and Exposition*, 2018, 1-11.

Torgo, L. *Data Mining with R: Learning and Case Studies*, Chapman & Hall/CRC, 2011.

Wu, C., Mai, F. and Yu, Y., "Teaching Data Mining to Business Undergraduate Students Using R", *Business Education Innovation Journal*, 7(2), December 2015, 64-73.

Weiss, S.M., Indurkhya, N. and Zhang, T., *Fundamentals of Predictive Text Mining*, Second Edition, Springer-Verag London, 2015.

Witten, I. H., Frank, E. and Hall, M. A., *Data Mining: Practical Machine Learning*

Tools and Techniques, 3rd edition, Morgan Kaufmann (Elsevier), 2011.

Zhao, J. and Zhao, S., "Business analytics programs offered by AACSB-accredited U.S. colleges of business: A web mining study", *Journal of Education in Business*, 91(6), 2016, 327-337.