Data Mining Course in the Undergraduate Computer Science Curriculum

Dr. Reza Sanati-Mehrizy, Utah Valley University

Reza Sanati-Mehrizy is a professor of Computer Science Department at Utah Valley University, Orem, Utah. He received his M.S. and Ph.D. in Computer Science from the University of Oklahoma, Norman, Oklahoma. His research focuses on diverse areas such as: Database Design, Data Structures, Artificial Intelligence, Robotics, Computer Aided Manufacturing, Data Mining, Data Warehousing, and Machine Learning.

Kailee Parkinson, Utah Valley University

I am a student at Utah Valley University majoring in Computer Science. I am currently an Engineering Writing Fellow and have written previously on implementing Data Mining courses at an undergraduate level. I am the recipient of the Student Excellence Award in Computer Science in the UVU College of Technology & Computing.

Dr. Elham Vaziripour, Utah Valley University

Elham Vaziripour, Ph.D. in computer science, is currently a professor assistant at Utah Valley University. Her area of research is Security, UX research, and Data analysis. She graduated recently, Dec 2018, from Brigham Young University. Her dissertation was on analyzing security and privacy of secure messaging applications.

Dr. Afsaneh Minaie, Utah Valley University

Afsaneh Minaie is a Professor and Chair of Engineering Department at Utah Valley University. She received her B.S., M.S., and Ph.D. all in Electrical Engineering from University of Oklahoma. Her research interests include gender issues in the academic sciences and engineering fields, Embedded Systems Design, Mobile Computing, Wireless Sensor Networks, Nanotechnology, Data Mining and Databases.

Data Mining Course in Undergraduate Computer Science Curriculum

Abstract

Data Mining combines tools from statistics, neural networks, and machine learning with database management to analyze large data sets. It is a well-researched area of computer science with high demand due to its usefulness in any field with large quantities of data where meaningful patterns and rules can be extracted. Therefore, many organizations and businesses can benefit from Data Mining techniques as these organizations record a massive amount of data daily.

The field of Data Mining is growing rapidly and there is increasing interest in providing students with a foundation in this area. It is crucial that the emerging field of Data Mining be integrated into the Computer Science curricula.

This paper will study different approaches that are used by different institutions of higher education to integrate Data Mining concepts into their curriculum. We use this data to make recommendations of how it should be taught in the undergraduate computer science program and what to include in course(s) in general and specifically in our institution.

Introduction

Data mining involves analyzing large data sets and also involves tools from Statistics and Artificial Intelligence, like Neural Networks and Machine Learning. The ability to use these tools to assess data is in increasing demand by employers. There is a vast amount of data available to businesses. These data are required to be assessed and analyzed for gaining insight in business decisions, behavioral studies, consumer habits, and many more areas of application. Other fields of study besides computer science have had an increasing interest in data science. Current listings in the realms of physics, biology, medicine, and advertising jobs indicate that companies are looking to hire individuals with the ability to analyze and extract useful patterns from data in addition to their academic field of study.

An adequate curriculum for Data Mining undergraduate courses needs to be created in response to an increasing demand for graduates to have the Data Science knowledge, skills, and ability to utilize and apply the necessary analysis tools. Our paper attempts to resolve the gap in the current state of unstandardized Data Science curriculum and presents a 15-week Introduction to Data Mining course geared towards undergraduate level students. This includes recommendations for programs and textbooks based on our research.

There are many reputable colleges that offer Data Science courses as part of their Computer Science degree. Also, there are many online teaching platforms dedicated to this subject. However, there is no clear consensus on what knowledge students need in order to be able to accomplish the tasks that employers are expecting. We are in the process of developing a comprehensive curriculum tailored to teaching students the necessary Data Mining theory and skills in our institution.

Our proposed curriculum is influenced by data gathered from syllabi of courses taught in the area of Data Mining at different institutions of higher education across the United States. Each syllabus was used to gather information for the topics taught in the class, the textbook and reading materials, the primary language used to teach students the concepts, and the frequency of homework assignments, tests, and projects. We used this data to weigh which topics are more heavily favored within academia. Then we used Data Science textbooks to build a feasible curriculum for Data Mining course in this paper.

Related work

There have been a few studies on different approaches that have been taken by online and offline training platforms. These studies indicate that there is no consensus on the syllabus. In some cases, there is a high overlap between materials that have been taught in the and machine learning area, which resulted in the field of data science emerging. This work is going to be a precious asset for the schools like UVU that have not set up the Data Mining course yet and are in the process of developing their program.

Ganesh [5] discusses how Data Mining is not just for computer science and should be offered in the field of statistics. Romero and Ventura [2] studied an interdisciplinary field of educational Data Mining (EDM) and reviewed milestones, applications, tools and future insights within this field. Anderson et. al. proposed a four-year undergraduate program in predictive analytics, machine learning, and Data Mining implemented at the College of Charleston [4]. However, they did not describe the details for any of these courses. Sanati-Mehrizy et. al., studied different approaches that have been taken by different institutions for integration of Data Mining concepts into undergraduate computer science program [1]. Their work motivated us to conduct our study on a larger scale. We also recommend a curriculum for undergraduate course of Data Mining. Our proposed curriculum has been inspired with the work by Chakrabarti et. al. [3], but we are considering new concepts such as big data analysis in this study.

The Rationale for Teaching Data Mining

To justify including Data Mining courses in undergraduate Computer Science programs we conducted an investigation into fields of study that have demand for graduates to have data science skills and experience. We chose six different majors listed below and searched for jobs related to that field on employment websites like Indeed, Ziprecruiter, and Workopolis. For each major we found ten different job offerings that required a degree in the field of study and for the applicant to have some knowledge of Data Science principles. Each job listing selected for our data set required a degree in the field of study, as well as listed data science skills. Key phrases that were frequently used in our data set emphasized: data analysis, data visualization, ability to analyze data, etc. The key data points we collected for analysis include:

Field of study:

Physics, Biology, Marketing, Medicine, Electrical Engineering, and Computer Science

Collected data points:

Job titles, company titles, listed Data Science requirements

Our collected data show that there is a general demand for Data Mining knowledge that makes the graduated students prepared for the highly competitive job market.

Methodology

We collected university curriculum data from a sample of colleges across the United States that offer Data Mining, artificial intelligence, or machine learning in their undergraduate programs. These universities were either selected from the sampled universities in the work by Sanati et. al. [1], or from other universities in this field [6]. We used the Google search engine to traverse through course description pages offered by the universities, online available syllabuses, and posted teaching materials. The university sample set and collected data points include:

University Sample Set:

University of Utah, Brigham Young University, Purdue University, Columbia University, New York University, Carnegie Mellon University, North Carolina University, Michigan State University, Illinois Institute of Technology, Washington University in St. Louis, University of Illinois

Collected Data Points:

Prerequisites to the course, main topics covered, subtopics included, referenced materials, programs utilized, assessment standards

Since Data Science courses offered on online platforms reflect the current expectations of the industry, we've included a small sample of Data Science courses offered by *Coursera* and *fast.ai.* Our approach for gathering data from these courses was the same as our approach towards universities to keep our data consistent.

The data points mentioned above were then used to create a 15 weeks Data Mining course curriculum. Topics were included in the curriculum based on the frequency they occured in our data, indicating that this topic was covered in multiple courses. We also considered the feedback from faculties who are currently teaching in the field. The course content that we considered, has been mostly covered with Data Science textbooks, such as *Principles of Data Mining* [7] and *Data Mining: Practical machine learning tools and techniques* [20].

Results

We present our collected data in the following sections.

- **Sample Universities**

Table I presents the raw data of sampled university courses that we collected for our study. This table displays the stated prerequisites of the course and the methods of assessment for the universities.

TABLE I - University Data Mining course details

University	Course Name	Prerequisites	Assessment
University of Utah [8]	Data Mining	Probability, Linear Algebra, Basic Programming, Data Structures	Unknown
Brigham Young University [9]	Intro to Machine Learning and Data Mining	Fundamental Algorithms, Data Structures, Complexity Classes	Unknown
Purdue University [10]	Introduction to Data Mining	Database Programming, Reasonable programming background, Discrete Math and Calculus	Programming Projects, Analytical written homework problems, Midterm and Final, Final Project, Open note/open book exams
Columbia University [11]	Machine Learning	Multivariate calculus, Linear Algebra, Basic Probability, Comfortable processing and analyzing in Python, Basic Algorithm Design and Analysis, Math	Two in-class exams, both comprehensive, 4-6 homeworks, half before the exam
New York University [12]	Applied Data Analytics	Basic Python	Semester-long group Project
Carnegie Mellon University [13]	Data Mining	Unknown	5 weekly homework assignments,

			Lab requirement, Two exams, Final project
North Carolina University [14]	Automated Learning and Data Analysis	Discrete Math, LOG 201 Logic, Probability and Statistics for Engineers, Introductory Linear Algebra and Matrices	Midterm and final, four homework assignments and semester long project
North Carolina University [15]	Artificial Intelligence	Data structures, Discrete math, Symbolic logic	Two to three programming assignments, one to two alternate assignments, Midterm and Final
Michigan State University [16]	Big Data Analysis	Basics of Algorithms and Data Structures, Basic Programming	Homework exercises, weekly programming assignments, Final Project
Illinois Institute of Technology [17]	Data Mining	Unknown	Midterm and Final
Washington University in St. Louis [18]	Data Mining	Programming ability	Quiz in class for each topic, three homework assignments, No exam, Project and report
University of Illinois [19]	An Introduction to Data Warehousing and Data Mining	Data structure and software principles, Programming ability	Unknown

The data from the above table shows that about 42% of the sampled university courses explicitly required students to have a background in mathematics. Of that subset of approximately five courses, three explicitly stated that students should have exposure to probability, three required linear algebra, and three expected understanding of discrete math. However, only two courses required students to have adequate experience in Calculus. The remaining courses in the entire sample either did not require mathematics or did not specify.

The courses varied in their approach to assess students, but some common themes arose between them. Four courses required students to work on a semester long project. 8 of the 12 courses assigned programming and theory homework, but varied greatly in the implementation, with some assigning weekly projects and some only requiring between two to six homework assignments for the semester. This also correlates with the assumption that students have an adequate programming background. This is supported by a majority of the classes, 75%, explicitly requiring programming skills as a prerequisite to the course.

Table II provides more information about the types of programs that students would be required to understand and work within for the duration of the class.

TABLE II - University Data Mining course resources

University	**Course Name**	**References**	**Programs**
University of Utah [8]	Data Mining	Mathematical Foundation for Data Analysis by Professor; Mining Massive Data Sets by Anand Rajaraman, Jure Leskovec, and Jeff Ullman; Foundations of Data Science by Avrim Blum, John Hopcroft and Ravindran Kannan	Python, Matlab/Octave
Brigham Young University [9]	Intro to Machine Learning and Data Mining	Unknown	Unknown

Purdue University [10]	Introduction to Data Mining	Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber; Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations by Ian H. Witten and Eibe Frank	WEKA
Columbia University [11]	Machine Learning	Pattern Classification by Duda, Hart, and Stork; A Course in Machine Learning by Daume; Convex Optimization by Boyd and Vandenberghe; Boosting: Foundations and Algorithms by Schapire and Freund	Unknown
New York University [12]	Applied Data Analytics	Big Data and Social Science: A practical guide to models and tools by Taylor Francis , Ian Foster, Rayid Ghani, Ron Jarmin, Frauke Kreuter	Python, SQL, Github

		and Julia Lane	
Carnegie Mellon University [13]	Data Mining	An Introduction to Statistical Learning: with Applications in R by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani Data Mining: Practical Machine Learning Tools and Techniques by Witten and Frank Elements of Statistical Learning by Hastie, Tibshirani, Friedman Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking by Provost and Fawcett Applied Predictive Modeling by Kuhn and Johnson	R
North Carolina University [14]	Automated Learning and Data Analysis	Introduction to Data Mining	R and Matlab

		by Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Tan Pang- Ning, Steinbach Michael Paperback	
North Carolina University [15]	Artificial Intelligence	Artificial Intelligence: A Modern Approach by Stuart Russell and Peter Norvig	Java
Michigan State University [16]	Big Data Analysis	Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar Mining of Massive Datasets by Anand Rajaraman, Jeff Ullman Hadoop: The definitive guide by Tom White Python for Data Analysis: Data Wrangling with Pandas, Numpys, and iPython by Wes McKinney	Java, Python, SQL
Illinois Institute of Technology [17]	Data Mining	Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne,and	Unknown

		Vipin Kumar	
Washington University in St. Louis [18]	Data Mining	Data Mining: Practical Machine Learning Tools and Techniques by I.H. Witten, E. Frank and M.A. Hall Data Mining: Concepts and Techniques by J. Han, M. Kamber and J. Pei	Unknown
University of Illinois [19]	An Introduction to Data Warehousing and Data Mining	Data Mining: Concepts and Techniques by J. Han, M. Kamber and J. Pei	Unknown

As demonstrated in Table II, there are many resources available for teaching the course of Data Mining. The sampled university courses vary in the materials and the amount of material they include to teach the course. However, there were three textbooks that appeared the most frequently within our sample set. Each of these textbooks was referenced for 3 of the 12 sample courses analyzed:

- *Data Mining: Practical Machine Learning Tools and Techniques* by I.H Witten, E. Frank and M.A. Hall
- *Introduction to Data Mining* by P. Tan, M. Steinbach, A.Karpatne and V. Kumar
- *Data Mining: Concepts and Techniques* by J. Han, M. Kamber and J. Pei

These textbooks are the most popular among our sample set and can be included as course textbooks.

Twelve universities required students work with a specific programming languages. The following list shows their frequencies.

- Python - 3
- Matlab - 2
- R - 2
- SQL - 2

- Java - 2
- WEKA - 1

➢ **Sample Online Platforms**

Table III presents information gathered from the online course curriculums, in the same format as the university courses given above.

TABLE III - Online Data Science course details

Online Platform	Course Name	Prerequisites	Assessment	Programs
Coursera [21]	Data Science Specialization	None	Online quizzes, Final Project	R, RStudio, Github
Fast.ai [22]	Practical Deep Coding for Learners, pt 1	Year of Coding Experience, High School Math	None	Python
Fast.ai [23]	Cutting Edge Deep Learning for Coders, pt 2	Practical Deep Coding for Learners pt 1, CNNS (resnets), RNNS (LSTM & GRU), SGD/Adam, Batch normalization, Data augmentation, PyTorch, numpy	None	Python

The teaching of online courses is very different from in-person university courses, so our main focus was to get a sense of the material these courses covered. However, the raw data that is displayed above still gives insight for constructing undergraduate Data Science courses. This table reveals that these classes have much lower prerequisite requirements, and very little strategy of assessment, if any. These courses may be useful in guiding curriculums for students who do not have a background in Computer Science but still have an interest in applying Data Science theory and skills within their field of study. This table also shows the popularity of Python and R languages in such online courses.

- **Major Themes and Topics**

Table IV presents key major topics and subtopics that were covered in the Data Mining courses from our sampled Universities. This table does not include outliers, such as topics that were only covered in one or two courses, and instead gives a basic overview of what types of topics are typically expected within a Data Mining curriculum.

TABLE IV - Topics and subtopics covered In course curriculums

Topic	Subtopics
Statistics	Probability, Subjective Probability, and Conditional Probability, Expected Values, Variability, Distributions, Asymptotics, Confidence Intervals, Hypothesis Testing, P-values, Resampling, Bayesian Estimator, Statistical Modeling, Independence
Clustering	Hierarchal, K-means, and Spectral Clustering, Anomaly Detection, Nearest Neighbor Methods, Biclustering, Probabilistic and Spectral Clustering
Regression	L1 Regression and Lasso,Linear Regression, Simple Logistic Regression,Least Squares, Multivariable Regression,Poisson Regression, Regularized Regression,Polynomial Regression, Model selection and validation in Regression, Local Regression,Logistic Regression decision boundary,Principal Components Regression
Big Data	Data Warehousing, Inference and errors associated, Bias/Variance
Classification	Tree based approaches, Neural Networks, Bayes Classifier, Classification modeling, Discriminative Classifications
Ethics	Privacy, Security, Data Leaks, Biases, Confidentiality, Societal consequences
Error measure	Evaluation of Models, Confusion Matrices, ROC Graph
Association Rules	Association Rule Mining, Patterns in frequent set,

	Efficient algorithms for finding frequent set, Association Analysis

Discussion

Based on results of our study, we built our curriculum around the idea that students would have adequate background in statistics and probability theory and a decent background in programming. Our curriculum includes a review of relevant statistics and probability concepts, but otherwise assumes that student's programming ability is relatively complete. Our proposed curriculum recommends the following for student assessment:

- Midterm and Final exams
- Small Lab Project for each topic
- A general research (possibly group) project for the semester

For our proposed curriculum, we recommend using Python, and specifically utilizing Anaconda tools and orange library, which makes it easier to work with normal data. For the big data requirements, we propose dedicating a week towards MapReduce and Hadoop.

Table V presents our proposed curriculum for an undergraduate level Data Mining course. The topics presented require students to fulfill the following prerequisites:

- Ability to program in at least one language
- Exposure to statistics concepts
- Understanding of Data Structures and Algorithms
- Understanding of Database theory

We choose to review Statistics concepts in Week 3, but require students to have had exposure previous so that teaching is not too comprehensive. The curriculum will require students to interact with databases in order to gather data. Therefore, an understanding of databases and data structures will be vital for students to engage with the class.

TABLE V - Recommended curriculum

Week	Main Topics	Topic Details
Week 1	Background and overview	What is machine learning (supervised vs. unsupervised methods), Databases, Intro to the data languages covered, Possible projects, Introduce tools(R, Python, Weka)
Week 2	Data Preparation	Collecting data(xml,json),

		Storing data,Cleaning data, Manage missing data,introduce some, repositories(UCI,Kaggle)
Week 3	Review statistic principles	Probability, Statistics, Naive Bayes learning methods, Relationship of statistics to machine learning
Week 4	Introduction to classification	Linear Regression, Naive Bayes
Week 5	Introduction to classification	Nearest Neighbor Classifiers
Week 6	Evaluate the accuracy of classifiers	Separate Training and Test sets, Cross Validation, Confusion Matrices,ROC graphs and curves
Week 7	Data Visualisation	Data Exploration and Visualization
Week 8	Decision Tree	Random Forest
Week 9	Avoid overfitting of Decision Tree, and Bias	
Week 10	Support Vector Machines	
Week 11	Clustering	K-means clustering, Hierarchical clustering
Week 12	Correlation, Causation	Association rule mining
Week 13	Applications	Natural language processing, Text mining
Week 14	Convolutional neural network	Deep learning
Week 15	Big data analysis	Hadoop, MapReduce

Week 1 through Week 3 are focused on introducing the students to basic Data Mining concepts. Week 1 focuses on giving background and context to the information being taught to students. Students will get basic exposure to handling data by Week 2. Week 3 is dedicated to reviewing statistics principles. These weeks are meant to reinforce principles and build off of basic knowledge that students already know and introduce them in a new context.

Week 4 through Week 6 introduce new theories and applications. Week 4 introduces Classification and Regression, and Week 5 introduces Nearest Neighbor Classifiers. Week 6 is focused on verifying accuracy and tools that students can use to engage with their data to ensure its quality. This gives the student concrete knowledge within the field of Data Mining.

The next weeks cover some basic skills and introduce common concepts within Data Mining theory. Week 7 is dedicated to the visualization of data and gives students the opportunity to build skills upon the basic concepts covered in the second week. Week 8 and Week 9 are dedicated to Decision Trees and covers their extension into Random Forests, and introduces students to bias and how it can arise in their data. Week 10 focuses on Support Vector Machines.

Week 11 and Week 12 introduce students to more basic theories and common problems that arise in the Data Mining field. Week 11 covers clustering and some basic approaches that can be applied to the data. Week 12 defines correlation and causation, and introduces the student to Association Rule Mining. This helps students evaluate the association among the data that they collect.

The final weeks, Week 13 through Week 15, focus on real-world applications and potential topics that will be required of students in the workplace. Week 13 focuses on introducing students to applications of Data Mining and where it is commonly practiced. Week 14 drops into some deeper theory again and teaches students Deep Learning and Neural Networks. Week 15 covers Big Data analysis and gives an introduction to Hadoop and MapReduce, and their applications and limitations.

Note that the proposed curriculum is a very comprehensive schedule. Each university may adjust it to fit the needs of their students.

Conclusion and Future Work

The field of Data Mining is growing rapidly and there is increasing interest in providing students with a foundation in this area. It is crucial that the emerging field of Data Mining be integrated into the Computer Science curricula.

This paper studied different approaches that are used by various institutions of higher education to integrate Data Mining concepts into their curriculum. We employed this data to build a teaching model for the Data Mining course for our and similar institutions.

The proposed teaching model in this paper provides a 15 weeks curriculum that gives students a comprehensive understanding of Data Mining principles based on major concepts covered in academia and required by job market. The concepts included in the course are intended to prepare students for careers that involve applying Data Mining skills.

We do not have adequate evidence and feedback from the students who have taken the course to verify that the provided strategies are the best overall approach. Future study may include a statistical analysis of how effective our proposed curriculum is in accomplishing our goals of preparing students with Data Mining skills, catering to undergraduate level students, and presenting concepts relevant to Data Mining. This would solidify our proposed curriculum's effectiveness and provide valuable feedback to make updates.

References

[1]. Sanati-Mehrizy, Reza, Kailee Parkinson, and Afsaneh Minaie. "Integration of data mining course in computer science curriculum." *Journal of Computing Sciences in Colleges* 34.2 (2018): 87-98

[2]. Romero, Cristobal, and Sebastian Ventura. "Data mining in education." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 3.1 (2013): 12-27.

[3]. Chakrabarti, Soumen, et al. "Data mining curriculum: A proposal (Version 1.0)." *Intensive Working Group of ACM SIGKDD Curriculum Committee* 140 (2006).

[4]. Anderson, Paul, et al. "An undergraduate degree in data science: curriculum and a decade of implementation experience." *Proceedings of the 45th ACM technical symposium on Computer science education*. ACM, 2014.

[5]. Ganesh, Siva. "Data mining: Should it be included in the statistics curriculum?." *The 6th international conference on teaching statistics (ICOTS-6), Cape Town, South Africa*. 2002.

[6]. Jain, Aarshay. "10 Analytics/Data Science Masters Program by Top Universities in the US" *Analytics Vidhya.* July 10, 2016. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/07/10-analytics-data-science-top-universities-masters-usa/. [Accessed Jan. 18, 2019].

[7]. Bramer, M, *Principles of Data Mining.* Springer, London: Springer-Verlag London Ltd., 2016.

[8]. University of Utah. *Data Mining.* http://www.cs.utah.edu/~jeffp/teaching/cs5140.html. [Accessed Jan 9, 2019].

[9]. Brigham Young University. *Introduction to Machine Learning and Data Mining.* https://cs.byu.edu/node/19963/47548. [Accessed Jan 9, 2018].

[10]. Purdue University. *Introduction to Data Mining.* https://www.cs.purdue.edu/homes/clifton/cs490d/. [Accessed Jan 12, 2019].

[11]. Columbia University. *Machine Learning.* http://www.cs.columbia.edu/~djhsu/coms4771-f18/. [Accessed Jan 12, 2019].

[12]. New York University. *Applied Data Analytics.* https://steinhardt.nyu.edu/scmsAdmin/media/users/ms184/Syllabus-AppliedDataAnalytics_June22_v1.pdf. [Accessed Jan 20, 2019].

[13]. Carnegie Mellon University. *Data Mining.* http://www.andrew.cmu.edu/user/achoulde/95791/. [Accessed Jan 20, 2019].

[14]. North Carolina University. *Automated Learning and Data Analysis.* https://www.engineeringonline.ncsu.edu/course/csc-522-automated-learning-and-data-analysis-2/. [Accessed Jan 20, 2019].
[15]. North Carolina University. *Artificial Intelligence I.* https://www.engineeringonline.ncsu.edu/course/csc-522-automated-learning-and-data-analysis-2/. [Accessed Jan 20, 2019].
[16]. Michigan State University. *Big Data Analysis.* https://www.engineeringonline.ncsu.edu/course/csc-522-automated-learning-and-data-analysis-2/. [Accessed Jan 20, 2019].
[17]. Illinois Institute of Tech. *Data Mining.* www.cs.iit.edu/~abet/CourseOverviews/cs422.doc. [Accessed Jan 20, 2019].
[18]. Washington University in St. Louis. *Data Mining.* https://www.cse.wustl.edu/~zhang/teaching/cs514/sp15/index.html. [Accessed Jan 20, 2019].
[19]. University of Illinois. *An Introduction to Data Warehousing and Data Mining.* https://wiki.illinois.edu//wiki/display/cs412/2.+Course+Syllabus+and+Schedule. [Accessed Jan 20, 2019].
[20]. Witten, Ian H., et al. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
[21]. Coursera. *Data Science Specialization.* https://www.coursera.org/learn/data-scientists-tools. [Accessed Jan 18, 2019].
[22]. Fast.ai. *Practical Deep Learning for Coders, pt 1*. http://course18.fast.ai/index.html. [Accessed Jan 18, 2019].
[23]. Fast.ai. *Cutting Edge Deep Learners For Coders, pt 2*. http://course18.fast.ai/part2.html. [Accessed Jan 18, 2019].