
A MODEL FOR AUTOMATED MATCHING BETWEEN JOB MARKET DEMAND AND UNIVERSITY CURRICULA OFFER

Ylber Januzaj, Artan Luma, Besnik Selimi, Azir Aliu, Bujar Raufi, Halil
Snopçe

Faculty of Contemporary Sciences and Technologies, South East European
University

DOI: 10.1515/seeur-2017-0024

Abstract

Technology plays a very important role in virtually all areas, and has become an inseparable part of the industry. Currently, industry and technology are at a high point of development and research, but there is an ever increasing gap between the market needs and the skills that universities deliver to students. There is an increasing need for consolidation between university curricula and the industry needs in terms of qualifications. In this paper we will present a description of the current state of the labor market in the field of technology, including the needs that arise in improving the existing curricula of the Universities. We review the different technologies that can be used, in order to

automatically gather information about the market needs in terms of job offers, and how they can be compared against University curricula. We will also present the latest achievements on these methods, and the suggestions that the researchers provide.

Index Terms: Technology, Data-Mining, Job Market, University, Curricula.

Introduction

Technology plays a very important role in virtually all areas, and has become an inseparable part of the industry. Currently, industry and technology are at a high point of development and research, but there is an ever increasing gap between the market needs and the skills that universities deliver to students. There is an increasing need for consolidation between university curricula and the industry needs in terms of qualifications. Studying relevant program has a high importance for students and the labor market. Currently in the field of technology, there are many labor market demands that fail to be met for various reasons, one of which is the lack of professional preparation in the relevant field. Applying technology to the industry also has a high importance, as it directly affects the country's income, almost in all areas. In (Anicic, Divjak and Arbanas, 2017) we can observe that the importance of technology lies on the fact that 5% of Europe's Gross Domestic Product (GDP) is directly dependent on the field of technology. How technology is important, is shown in some research that has been done on trends and job requirements in this area for the period 2012-2020. In (Anicic, Divjak and Arbanas, 2017) it is mentioned that,

according to the "Main Forecast Scenario" from 2012, an increase of 7.4 million to 7.9 million job vacancies is expected in 2020, according to "Stagnation Scenario", an increase of up to 7.8 million job vacancies by 2020, and according to "Disruptive Boost", an increase of 8.1 million job vacancies is expected by 2020. Therefore, based on these forecasts in all three cases, which are claimed to be optimistic values, job vacancies in the field of Technology will exceed the number of graduates in this area, and at the same time we can notice the great importance of studying relevant programs in the field of Technology. Earlier, it was mentioned that there are many reasons that impact the requirements to be not sufficiently met by graduate students in the market, such as lack of experience, poor knowledge of international languages, and weaknesses in communication, as well as studying in the right direction based on labor market requirements. In (Aziz & Yusof, 2016) it has been mentioned that as far as it is important to complete the studies and get the status as a "graduate", it is also important to determine the quality of the program offered by the relevant institution. Based on the importance of such researches, the application of some sub-fields of technology is required in order to have a more accurate and optimistic result. Such research also requires a combination of the field of technology and mathematics, as besides Data Mining methods, such as classification, clustering, etc., also mathematical techniques such as Bayes, Regression, Decision Tree, descriptive statistics, etc. will be applied. Below we will present the latest achievements in this regard, as well as the recommendations that researchers provide for the future.

Overview

Currently there is a very large amount of data in the information industry. But these data, if not converted into useful information, remain as unnecessary data. It is imperative that these data be converted and extracted from the information they can use. It is not necessary to do only extract of these data, but we need also to do data cleaning, data integration, data transformation, data mining, and data presentation.

So, Data Mining can be defined as the process of extracting knowledge from the data. The knowledge that can be extracted and used in several areas such as:

- *Market Analysis.*
- *Education.*
- *Customer Retention.*
- *Production Control.*
- *Medical Analysis.*

So, as we can see, Data Mining can also be applied in the Education field. In our case we use it for classification and data clustering, and with the application of Data Mining we ensure that we have very effective and reliable classification and clustering. On the other hand, Machine Learning is a field that has enabled us to automate many processes that we use today to ease our work in many aspects. In our case we will apply Machine Learning in order to automate our system to be able to give recommendations.

Literature review

We apply Data mining techniques for clustering and matching algorithms for matching in order to implement our research. Also as a part of our research is web scraping and crawling from websites which offers informations about vacancies. In different papers it is mentioned that Research Questions accompany us during all the time of research, and as an important part first we need to define them. After the define Research Questions, then we give answers based on literature review. Our literature review is based in these Research Questions:

Is there any increasing need for consolidation between university curricula and the industry needs in terms of qualifications?

This question identifies the need of consolidation between University curricula and the market demands.

What is the level market demands that fail to be met by Universities?

This question shows the level of market demand that actually fails to be met by curricula which are offered by Universities in our region.

Are Higher Education Institutions (HEI) actually facing difficulties in improving and managing their work and organization?

This question shows if actually the HEI are facing difficulties in improving and managing their work and organization. Also it shows the techniques which are actually used by universities in order to manage and organize their work.

What is the limitation and biggest challenges of Data Mining related to matching labour market profiles and university curricula?

This question identifies which is the biggest challenge of Data Mining techniques related to the problem, and which are the steps to overcome these challenges.

Which content has become the target of all researchers of the academy?

This question shows target of all researches of the academy which actually is used.

Papers

We listed all Research Questions, and as we mentioned above they will accompany us during all the time of our research. In order to get answers for all these Research Questions we needed to read different articles from different sources. We analyzed in total of 78 papers from **IEEE, ACM, and Google Scholar**. Papers are published in different fields, and their publication year is between **1983** and **2017**. Next in Table 1 we show the separation of papers by Libraries.

Library	Number of Papers	Percentage
ACM	25	32%
IEEE	33	42%
Google Scholar	20	26%

Table 2. Publications by Libraries

As we can see in Table 1 we used three digital libraries in order to analyze the papers. Also above we can see that the most papers are used

from **IEEE** library which is the most appropriate library for the field of technology, also we have a considerable number of papers downloaded from **ACM** which also is an appropriate library for the field of technology.

Field of Interest	Number of Papers	Percentage
University Curricula	30	38%
Data Mining	30	38%
Machine Learning	10	13%
Web Crawler / Scraping	8	11%

Table 3. Publications by Field of Interest

As we can see in Table 2 analyzed papers are separated in four fields:

University Curricula, Data Mining, Machine Learning and Web Crawler / Scraping.

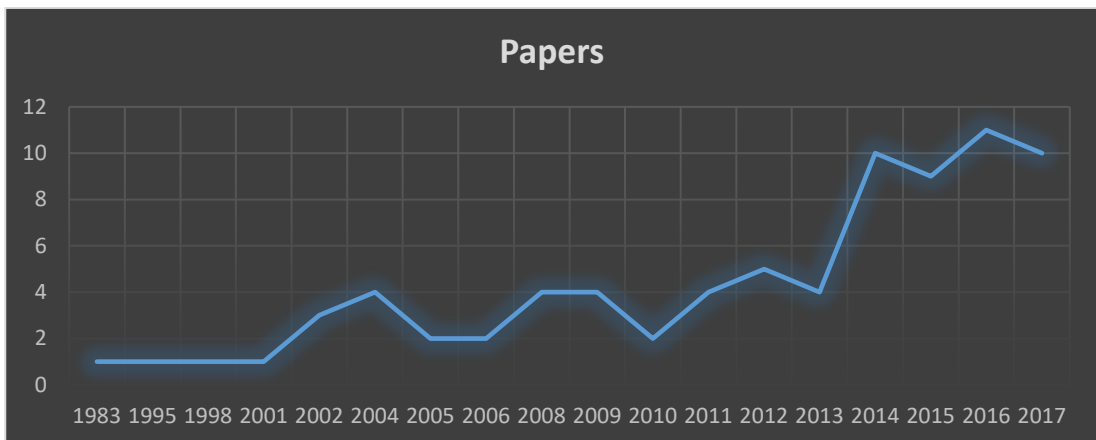


Figure 9. Publications over the Years

As we can see in Figure 1 the publication year of the papers are between **1983** and **2017**. There are some basic papers that we had to analyze even they were old. But as we can see the most papers are published in the last three years.

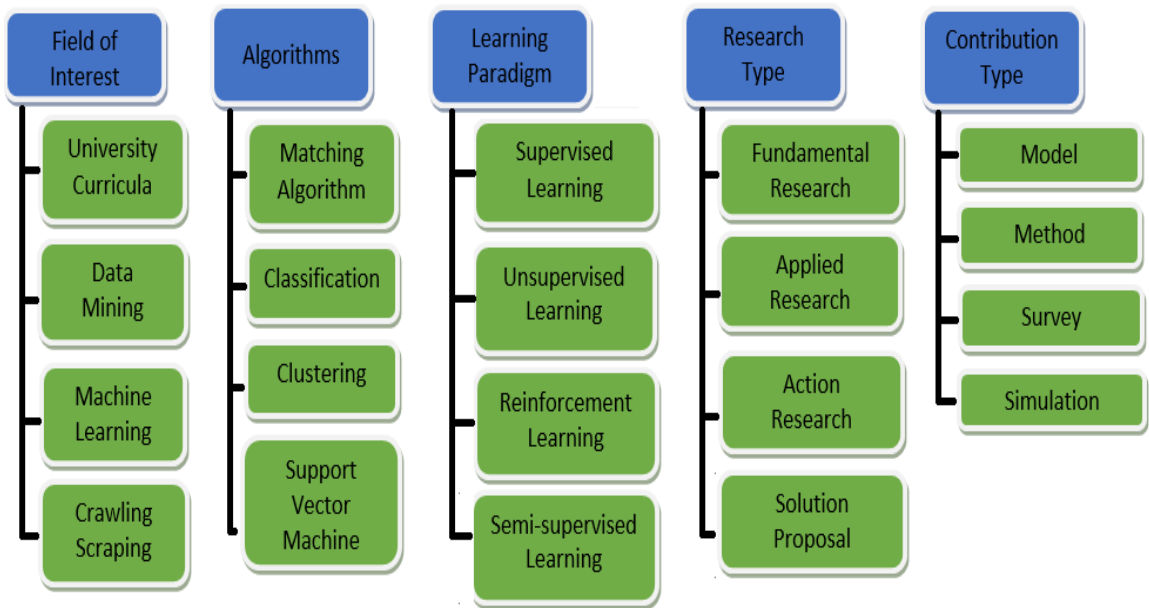


Figure 10. Classification scheme

State of the Art

Nowadays the technology has reached a high point of development and research, and although WWW technology is found in its third decade, the application of Data Mining on websites remains one of the biggest challenges (Thakar, Mehta and Manisha, 2015). It is precisely the achievement and dynamics of the Web that has made the application of Data Mining in websites more difficult. Therefore, a technique of extracting information from websites needs to be applied in our case. According to (Sahu & Bhatt, 2017) Data-Mining offers powerful analysis techniques for different areas, especially appropriate techniques for the education field. Seeing the development of the education field, data growth for students, then Data Mining remains one of the most perfect techniques for achieving a successful research. When we talk

about the field of education, then specific techniques are required for this field, as is the case with "Educational Data Mining" or EDM, where the main purpose is data processing in the field of education. Why an application of such a technique is important, it is shown by the increase of the data capacity in the education institutions' databases in our country and in the region, or otherwise known as Big Data. Just as Data - Mining works, EDM uses the same techniques such as: classification, clustering, and association rules.

a) The classification is used to classify the existing data based on the training set, and uses the template to classify a new type of data.

b) The clustering is used to divide into data sets that resemble each other, and those that differ from each other.

c) Association rules are used to detect links and dependencies between variables in large databases or Big Data.

In the third research question of the literature review we can see that Higher Education Institutions all over the world are facing difficulties in improving and managing their work and organization, this is also mentioned in (Sahu & Bhatt, 2017). Therefore, in order to achieve this goal, Data Mining is considered as one of the most appropriate techniques which help these institutions make better decisions in order to improve their activity. Therefore, during the literature research, we have noticed several different analyzes that have been made by the Universities, such as: student performance, student challenges with the real world, analysis of offered literature, classification of graduates, and so on. But there is still no research on the appropriateness of the programs offered by Universities and labor market requirements, and

such research is especially necessary in the countries of the region. In the fourth research question of the literature review we can see that one of the biggest challenges of Data Mining is data extraction within dynamic web sites. It remains one of the biggest challenges, since in order to apply Data - Mining within a web site, we need to undertake more depth analysis and research of web content. There are many factors like website dynamics, increased security, increasing data from seconds to seconds, and so on, which have caused this to be the biggest challenge. The application of data mining on the web is commonly referred as Web mining. According to (Bharanipriya & Prasad, 2011) web mining is nothing else, but just a data extraction from the web site by applying Data Mining techniques. Knowing that the content of the websites is not only textual, but there are also other data such as documents, tabular, unstructured data, and semi-structured data, etc. When working with such data, one should work with unstructured, or at best with what it can be considered as semi-structured data. It is precisely these two forms of data that we can hope for in the case of our research. In order to retrieve data from the web, one should consider different aspects such as: *Source finding, Data selection, Generalization, Analysis and Visualization*.

The **source finding** aspect includes the phase of determining the source from which the information will be extracted, whether this *on - line* or *off - line* source.

Once the source has been found, **data selection** is required, in which we need to find the right tools and definitions in order to extract meaningful data.

The **generalization** is also known as the assessment phase. According to (Bharanipriya & Prasad, 2011) at this stage, machine learning or data-mining processes are applied in order to identify general patterns within certain web sites.

At the stage of **analysis**, data accuracy is measured based on certain data parameters.

The last stage is the presentation phase or the stage of **visualization**. At this stage, it is decided what form the presentation of the processed data will be presented.

There are different techniques and tools that can be applied for mining web content. Some of these tools are based on extracting data from the HTML structure of web pages, possibly by using CSS selectors to focus at particular kind of data. Other tools may at first extract the text from web pages, and then extract concepts from row text. These techniques are commonly referred to as *web scraping*.

Once the content we are interested is extracted, it needs to be processed in order to obtain meaningful results. In the following we will talk about classification as one kind of processing, where we will present the latest achievements in this field.

The fifth research question of literature review shows that web content has become the target of all researchers of the academy, while it is also mentioned in (Thelwall, 2001). In order to handle the content of the webpages we need a special application, which will go through several phases. Also, processing textual content of web pages may be costly in terms of processing time (Thelwall, 2001). There are nothing else but

part of search engines in the form of an application that requires data in a certain, automated and regular way (Thelwall, 2001), (Adapure, Kale and Dharmik, 2014). Why crawlers are so important is the fact that search engines depend on them, enabling you to update information on up-to-date websites. Unlike web crawlers, we can encounter the terms: robots, mice, worms, pedestrians, etc., and it is as old as the internet, since the first crawler was created in 1993. The shape of a crawler is moving from page to page using the website structure, and downloading its content to locally store it in a particular destination. Below we will describe the web browser architecture of how it works.

Web crawler is divided into three components: frontier, downloader page and repository. The first component used to keep the list of unrecognized websites so far. Initially, the content of the existing link is extracted from the web site and downloaded, then checked for unrecognized links that are within the Web site. This link checking cycle continues until the frontier has been emptied. The second component or Page downloader is used to download web sites from the web that belong to links that are stored by the web crawler. As well as the third component or web repository is exactly where the content of downloaded websites is stored, but a large amount of data known as objects is stored here. The content that is stored is just HTML, all other types of data will be ignored. So these are the three components of how a web crawler works, but if we try to present the work of a web crawler through the steps then it looks like this:

1) Splitting visited and unvisited links, 2) Placement at the frontier, 3) Selecting links from the frontier, 4) Extraction of the web page content that is linked to

the boundary link, 5) Website analysis if there is any unrecognized link within it, 6) Placement of all unvisited links at the frontier, 7) Repeat step 2, until the frontier has been emptied.

So this is the work of a web crawler through the steps of how it works, and if we want to describe it works like this: We have the program starting point, then checking the visited and unlisted links. Check if there is a termination of the application, if yes it terminates, if not it continues and selects a url link from the frontier, if there is no link it is interrupted. If the link exists then the content of the web page is extracted and the web site content is checked if there is any unvisited link, if another link exists it is placed in the parser, and after this step we have a rewrite of the cycle known as the crawling loop.

With the growth and development of the web, Internet users find it harder to track and receive all relevant web information without any automated process (Haddaway, 2015). Currently there are many applications that enable web scraping, and based on (Grasso, Furche and Schallhart, 2013) current applications that offer web scraping are offered as ad-hoc using complicated tools and languages. There is also a great variety in how these applications are provided, for example: some of them are owned by someone else, and some of them are more likely to be implemented. According to researchers in this area, there is a great demand for such applications to be provided, since according to them, providers cannot expect to provide automated data extraction processes, since they are more concentrated in filtration and recommendations. In this paper we will present a project which is open source and is known as OXPath. This project is the continuation and expansion of XPath, with the difference

that it contains four functions more than the XPath previous version. Through this module we can do web scraping in a very fast and efficient manner and with less cost, also the users of this module can experiment with the existing Firebug and FirePath tools (Furche, Gottlob, Grasso, Schallhart and Sellers, 2012).

On other hand we can see that the most important unsupervised learning problem is considered the Clustering. We can find many definitions about Clustering, but the most appropriate definition is “organization of objects into groups which has similar members in some way”.

As a conclusion of literature review, we can see that all these tools are needed in order to create such a model. Also we can see that there is a need of creating an automated model for profile – curricula matching, which is our focus. So initially as a gap is identified: **Lack of automatized profile matching**. So figuring out the drawbacks that currently has matching profiles with curricula, where up to now we have been working manually in this direction and has cost over time, our model will be an automated model that will make automated profile matching. As sub - gap is identified: **Lack of automated design of reality gap**. So, as we know, currently Design Reality Gap is applied only manually, we will present it in an automated form to make a later comparison between the actual and the automated model.

Research Objectives

Our main idea is “Automated model for comparison of labor market and University curricula’s”. While this is our target we also emphasize the importance of the research:

1. *Universities need to adjust their Curricula to labor market.*
2. *Labor market demand profile “overflow”.*

Now in one side we have the main idea, and in other side we have the importance of the research. Based on this now we are able to define the problem which is known as the main gap of the research:

Are we able to automate the process of profile – curricula matches?

The idea is to make an automated comparison between market demands and university study programs by applying Data Mining techniques. First we scrap information from job vacancies web sites. How we do this? Actually there are web sites which publish job vacancies in different fields, our focus will be in the field of technology. The Information that will be extracted are:

- *Title of the vacancy.*
- *Position.*
- *Skills that are required.*
- *Advantages.*

These informations will be processed and will be clustered in order to create different types of profiles.

In other side we have the information about University curricula that will be extracted. While Universities in our region allow public access on study programs, these information will be automatically extracted. The information that will be extracted are:

- *Title of study program.*
- *List of subjects M or E.*
- *Credits of subjects.*

- *Number of hours for Lecture/Exercises.*
- *Description.*
- *Career.*
- *Learning Outcomes.*

Also these informations will be processed and will be clustered in order to create different types of profiles. After we have groups of profiles in both sides, we apply matching algorithm in order to compare the adjustment between curricula's to labor market. Next in Figure 3 we show a sketch of our proposed approach.

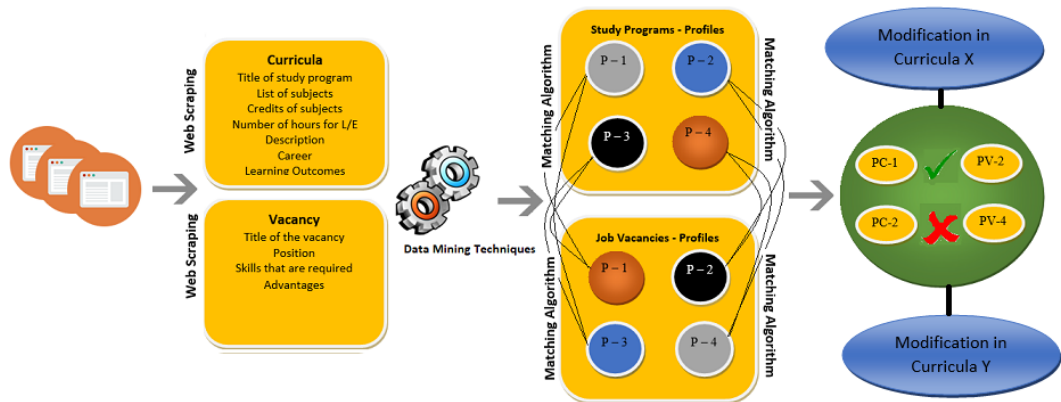


Figure 11. Sketch of proposed approach

In Figure 3 we can see the sketch of our proposed automated model for profile – curricula matching. First we have the scraping process from web sites, here we extract the content for the curricula and for vacancies. After we extract the data, now we apply Data Mining Techniques and we get different groups of profiles. After we have profiles now then we apply matching algorithms and we get matching between Profile Curricula “PC” and Profile Vacancies “PV”. Finally after we get the

matching between PC and PV, our model is able to give recommendation for any kind of modification in Curricula X or Y. These kind of modification includes removing a subject or adding a new one.

Hypotheses

1. *Automated methods can help matching curricula with workforce demand.*
2. *Curricula offer and job demands can be “quantified” using Data Mining Techniques.*
3. *An automated profile – curricula model will increase the level of Triple Helix cooperation.*

Research Questions

1. *What are the requests and recommendations in cooperation between market and universities?*
2. *Will automated profile – curricula matching decrease the adjusting time of companies?*
3. *How Universities are able to adapt to the recommendations that automated model exposes.*
4. *What is the current cooperation between universities – market in order to improve the quality of skills and their match with labor market demands?*

Methodology

The main purpose of this research is the suitability of university syllabus with the market demand in the region, as there is a major disadvantage in this regard. As we can see in the literature review, the adaptability of university programs to market demands is a problem that has begun to be addressed and is also required to work harder in this regard. Our main hypothesis **H0** is "**Automated Profile - Curriculum Compatibility**

Will Reduce Company Adaptation Time", and in order to prove this then an automated model should be created which will compile the level of suitability between the university syllabus and market demands. In order to arrive at such a pattern we will follow the steps below:

- 1) *Data source identification.*
- 2) *Web Crawling*
 - a. *Curricula*
 - i. *Subject.*
 - ii. *Description.*
 - b. *Job Vacancies.*
- 3) *Profile Clustering.*
- 4) *Recommendation.*
- 5) *Evaluation of the system.*

In the following we will describe each point in order to provide more detailed information on the methodology for how to implement such a model.

Data source identification.

In order to do a profile - curricula matching we will use the official data provided by the universities. The data that will be used are data on the study programs currently offered by these universities. Since all universities are obliged to make their data public, and there are countries in the region where the government provides us with open access to data, then these data will be used to achieve the implementation of such a model. It is not important in what format we get the data, as different techniques will be applied on these data where their normalization will take place.

Crawling

The second step of our research is to get crawled on some appropriate Web sites. The reason why Web Crawling will be applied is to extract information that is published by some websites on job vacancies. These data will be extracted in certain keywords, where their clustering will also be based on them. Initially in the region there are web pages that contain data on university programs, these data also contain descriptions of the subject and description of the study program. Based on this information will be also the clustering of study programs that are offered by universities. These later clustering's will be used to match the market requirements. On the other hand, Crawling will also be applied on web sites that provide information on competitions offered by different companies. Also, the web crawling application will be made to specific keywords, knowing that each contestant has additional information on the specific position requests. So, each position that is required keeps information in addition to the required position, the applicant must have knowledge in several different areas such as: programming, databases, networking, etc. All of these descriptions of later positions will be used to gather different positions.

Profile Clustering

In order to achieve a successful implementation of such a model, it is necessary to initially cluster the study programs and the clustering of the positions required in the competition. Initially with regard to the study programs, using the case description in the directions will be clustered into several groups, such as programming, computer networks, databases, etc. Also on the other hand where the vacancy positions are

going to be clustered based on the description of the positions. How will this clustering become? Initially, each position in the competition has its own description, and by collecting keywords that contain positions in the competition then it will be possible to cluster the positions by dividing them into several groups. These groups of positions later will be used to match the study programs that are also divided into groups.

Recommendation

A very important part of our model is the part of the recommendation that it will be able to give. Once a comparison is made between the profile and the market demands for a particular field, then our model will be able to give a recommendation if there is a need for changes in the study program. When we say to make recommendations for changes, here include the proposal for introducing any new subject into the study program, modifying some or removing any subject that is useless for market demands.

Evolution of the System

The last step of our model is the evolution of the system, where it will be tested the model of what evolution will have. Certainly such a model will be designed to be applied in different regions later on and in different areas where testing of the degree of suitability of study programs and market demands in different fields will be carried out.

Contribution

Our purpose is to create an automated model that makes comparisons between profiles and curricula. Actually there are manually researches

made on comparisons between profiles and curricula, and that is considered the main contribution of our thesis.

Nowadays in our region there are accredited more than 500 study programs and still there is a gap between universities and industry. After this research universities will be able to use our recommendations in order to modify their curricula's.

Except universities, another contribution will be for the companies that request graduate students that have perception in: creativity, problem solving, communication, team work, etc.

The advantage of our model is that it will be applicable in different countries with different companies. Even though our research will be made in a specific country: **Macedonia, Kosovo or Albania**, it will not be limited only for a country.

Current Work

We are currently in the first phase of implementing a Web Scraping application that will extract the content of all competitions published in the field of technology. Initially, it will be possible to identify the region where the research will be conducted, and then identify the websites that publish information regarding competitions offered each day by different companies. The first phase of our research will be the publication of these results, to continue later on with their elaboration.

Conclusion

During our work we have presented the importance of studying relevant program based to market demands. We have noticed the great importance of the technology field and the requirements that are currently in this area. We have presented the total number of vacancies in the field of technology, where we could see that it is a much larger number compared to graduates all over the world. Also in our paper are identified the methods by which such research can be achieved including: Data Mining, Classification, Web Scraping and Web Crawler. For each of these we have presented the latest achievements in this area and the recommendations that researchers have provided in this area. We have noticed how Data - mining has managed to solve many of the real world problems, and especially the achievements it has in the field of education. We have also noticed how Data - mining can process unstructured and semi - structured data. Several solutions to problems with the classification in the field of education were presented, where we could see how the performance of the student is based on the achievements of the old students. While in the Web Scraping and Web Crawler fields we have seen how these two methods actually enable web content extraction irrespective of how dynamic the web site is and how much information the web site has. So considering the achievements in this area and the needs that are currently present in the field of education, we think that a hybrid application of these methods will help us significantly in our research.

References

- [1] M. Agaoglu, “Predicting Instructor Performance Using Data Mining Techniques in Higher Education”. In *IEEE Access*, May 2016.
- [2] T. Xie, Q. Zheng, W. Zhang, H. Qu, “Modeling and Predicting the Active Video – Viewing Time in a Large – Scale E – Learning System”. In *IEEE Access*, June 2017.
- [3] A. M. Njeru, M. S. Omar, S. Yi, “IoT for Capturing and Mastering Massive Data Online Learning Courses”. In *IEEE Computer Society, ICIS*, Wuhan, China, May 2017.
- [4] R. Heartfield, G. Loukas, D. Gan, “You are probably not the weakest link: Towards Practical Prediction of Susceptibility to Semantic Social Engineering Attacks”. In *IEEE Access*, October 2016.
- [5] E. J. Fortuny, D. Martens, “Active Learning – Based Pedagogical Rule Extraction”. In *IEEE Transaction on Neural Network and Learning Systems*, Vol. 26, No. 11, November 2015.
- [6] A. Mukhopadhyay, S. Bandyopadhyay, “A Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part I”. In *IEEE Transaction on Evolutionary Computation*, Vol. 18, No. 1, February 2014.
- [7] Zh. Song, A. Kusiak, “Optimization of Temporal Processes: A Model Predictive Control Approach”. In *IEEE Transaction on Evolutionary Computation*, Vol. 13, No. 1, February 2009.
- [8] S. Malgaonkar, S. Soral, Sh. Sumeet, T. Parekhji, “Study on Big Data Analytics Research Domain”. In *International Conference on Reliability, Infocom Technologies and Optimization ICRITO*, Noida, India, September 2016.
- [9] K. P. Anicic, B. Divjak, K. Arbanas, “Preparint ICT Graduates for Real – World Challenges: Results of a Meta – Analysis”. In *IEEE Transactions on Education*, Vol 60, No. 3, August 2017.
- [10] A. Haskova, D. V. Merode, “Professional Training in Embedded Systems and its Promotion”. In *IEEE Transacions on Education*, 2016.
- [11] S.C. Smith, W. K. Al-Assadi, J. Di, “Integrating Asynchronous Digital Design into the Computer Engineering Curriculum”. In *IEEE Transactions on Education*, Vol. 53, No. 3, August 2010.

- [12] M. D. Koretsky, D. Amatore, C. Barnes, Sh. Kimura, “Enhancement of Student Learning in Experimental Design Using a Virtual Laboratory”. In *IEEE Transactions on Education*, Vol. 51, No. 1, February 2008.
- [13] B. G. Member, V. S. Sheng, K. Y. Tay, W. Romano, Sh. Li, “Incremental Support Vector Learning for Ordinal Regression”. In *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, No. 7, July 2015.
- [14] J. Li, T. Zhang, W. Luo, J. Yang, X. T. Yuan, J. Zhang, “Sparseness Analysis in the Pretraining of Deep Neural Networks”. In *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 28, No. 6, June 2017.
- [15] Y. Qian, F. Li, J. Liang, B. Liu, Ch. Dang, “Space Structure and Clustering of Categorical Data”. In *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 27, No. 10, October 2016.
- [16] Y. Xiao, B. Liu, Zh. Hao, “A Maximum Margin Approach for Semisupervised Ordinal Regression Clustering”. In *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 27, No. 5, May 2016.
- [17] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, Sh. Li, “Incremental Support Vector Learning for Ordinal Regression”. In *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, No. 7, July 2015.
- [18] P. Navrat, L. Molnar, “Curricula Transformation in the Countries in Transition: An Experience from Slovakia”. In *IEEE Transactions on Education*, Vol. 41, No. 2, May 1998.
- [19] S. Nalintippayawong, K. Atcharyachanvanich, “IT Management Status in Public Higher Education Institutions in Thailand”. In *IEEE ICIS 2016*, June 26-29, 2016, Okayama, Japan.
- [20] J. I. Godino – Llorente, R. Fraile, J. C. Gonzales de Sante, V. Osma – Ruiz, N. Saenz – Lechon, ““Design for All in the Context of the Information Society”: Integration of a Specialist Course in a Generalist M.Sc. Program in Electrical and Electronics Engineering”. In *IEEE Transactions on Education*, Vol. 55, No. 1, February 2012.

- [21] M. Dolores Cano, “Students’ Involvement in Continuous Assessment Methodologies: A Case Study for a Distributed Information Systems Course”. In *IEEE Transactions on Education*, Vol. 54, No. 3, August 2011.
- [22] Y. He, Ch. Wang, Ch. Jiang, “Mining Coherent Topics with Pre-Learned Interest Knowledge in Twitter”. In *IEEE Access*, June 2017.
- [23] H. Pirkkalainen, J. P. P. Jokinen, J. M. Pawlowski, “Understanding Social OER Environments—A Quantitative Study on Factors Influencing the Motivation to Share and Collaborate”. In *IEEE Transactions on Learning Technologies*, Vol. 7, No. 4, October-December 2014.
- [24] G. Goth, “Network-Enabled Compulsory Education Getting Big Push”. In *IEEE Computer Society*, February 2009.
- [25] R. Mehmood, F. Alam, N. N. Albogami, I. Katib, A. Albeshri, S. M. Altowaijri, “UTiLearn: A Personalised Ubiquitous Teaching and Learning System for Smart Societies”. In *IEEE Access*, February 2017.
- [26] J. J. Guerrero, L. A. Guerrero, “A Virtual Repository of Learning Objects to Support Literacy of SEN Children”. In *IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje*, Vol. 10, No. 3, August 2015.
- [27] A. A. Choudhury, J. Rodriguez, “A New Curriculum in Fluid Mechanics for the Millennial Generation”. In *IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje*, Vol. 12, No. 1, February 2017.
- [28] A. Sethi, “Factors Responsible for Mismatch between Demand and Supply of Requisite Skill in India”. In *IJARIE-ISSN (O)-2395-4396*, Vol. 3, Issue 3, 2017.
- [29] L. Anastasiu, A. Anastasiu, M. Dumitran, C. Crizboi, A. Homaghi, M. N. Roman, “How to Align the University Curricula with the Market Demands by Developing Employability Skills in the Civil Engineering Sector”. In *Education Sciences*, doi: 10.3390/educsci7030074, July 2017.
- [30] K. P. Anicic, B. Divjak, K. Arbanas, “Preparing ICT Graduates for Real – World Challenges: Results of Meta – Analysis”. In *IEEE TRANSACTIONS ON EDUCATION*, Vol. 60, No. 3, August 2017.

- [31] M. T. R. A. Aziz, Y. Yusof “Graduates Employment Classification using Data Mining Approach”. In *Proceedings of the International Conference on Applied Science and Technology*, ICAST, 2016.
- [32] S. Sahu, M. Bhatt, “Big Data Classification of Student Result Prediction”. In *International Journal of Research In Science & Engineering*, Volume: 3 Issue: 2 March-April 2017.
- [33] V. Bharanipriya, V. Kamakshi Prasad, “Web Content Mining Tools: A Comparative Study”. In *International Journal of Information Technology and Knowledge Management*, Volume 4, No. 1, pp. 211-215, 2011.
- [34] P. Thakar, A. Mehta, Manisha, “Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue”. In *International Journal of Computer Applications (0975 – 8887)*, Volume 110 – No. 15, January 2015.
- [35] G. Grasso, T. Furche, Ch. Schallhart, “Effective Web Scraping with OXPath”. In *WWW 2013 Companion, Rio de Janeiro, Brazil. ACM 978-1-4503-2038-2/13/05, 2013.*
- [36] M. Thelwall, “A Web Crawler Design for Data Mining”. In *Journal of Information Science*, pp. 319–325, 2001.
- [37] T. V. Adapure, R. D. Kale, R. C. Dharmik, “Study of Web Crawler and its Different Types”. In *IOSR Journal of Computer Engineering*, Volume 16, Issue 1, PP 01-05, 2014.
- [38] F. Ahmad, N. H. Ismail, A. A. Aziz, “Using Classification Data Mining Techniques”. In *Applied Mathematical Sciences*, Vol. 9, pp. 6415 - 6426, no. 129, 2015.
- [39] D. Garcia-Saiz, M. Zorilla, “Comparing Classification Methods for Predicting Distance Students Performance”. In *JMLR: Workshop and Conference Proceedings 17*, pp. 26-32, 2011.
- [40] N. R. Haddaway, “The Use of Web-scraping Software in Searching for Grey Literature”. In *The Grey Journal*, Volume 11, 2015.
- [41] T. Furche, G. Gottlob, G. Grasso, Ch. Schallhart, A. Sellers, “XPath: A language for scalable data extraction, automation, and crawling on the deep web”. In *The VLDB Journal*, 2012.
- [42] L. Auria, R. A. Moro, “Support Vector Machines (SVM) as a Technique for Solvency Analysis”. In *German Institute for Economic*

Research, 2008.

- [43] M. Awad, L. Khan, F. Bastani, “An Effective Support Vector (SVM) Performance Using Hierarchical Clustering”. In *IEEE 24th International Conference on Tools with Artificial Intelligence*, 2004.
- [44] A Brief Introduction to Support Vector Machine (SVM). January 25, 2011.
- [45] ACM Recommendations for Computer Science Curricula, Volume I, 1983.
- [46] Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering
- A Volume of the Computing Curricula Series, IEEE, 2014.
- [47] Th. Iliou, Ch. N. Anagnostopoulos, M. Nerantzaki, “A Novel Machine Learning Data Preprocessing Method for Enhancing Classification Algorithms Performance”. In *16th EANN workshops*, ACM, Rhodes Island, Greece, 2015.
- [48] M. M. Yusof, R. Mohamed, N. Wahid, “Benchmark of Feature Selection Techniques with Machine Learning Algorithms for Cancer Datasets”. In *ICAIR and CACRE '16*, ACM, Kitakyushu, Japan, 2016.
- [49] D. Brandon, “TEACHING DATA ANALYTICS ACROSS THE COMPUTING CURRICULA *”. In *CCSC: Mid-South Conference*, 2015.
- [50] H. Hu, J. Li, A. Plank, H. Wang, G. Daggard, “A Comparative Study of Classification Methods for Microarray Data Analysis”. In *Proc. Fifth Australasian Data Mining Conference*, 2006.
- [51] H. Liu, X. Yin, J. Han, “An Efficient Multi-relational Naïve Bayesian Classifier Based on Semantic Relationship Graph”. In *MRDM'05*, ACM, Chicago, Illinois, USA, 2005.
- [52] M. HooshSadat, H. W. Samuel, S. Patel, “Fastest Association Rule Mining Algorithm Predictor (FARM-AP)”. In *C3S2E 11*, Montreal, QC, Canada, 2011.
- [53] H. Hu, J. Li, “Using Association Rules to Make Rule-based Classifiers Robust”. In *Australian Computer Society, Inc.*, ACM, 2005.
- [54] A. Sun, E. Lim, W. Ng, “Web Classification Using Support Vector Machine*”. In *WIDM'02*, ACM, Virginia, USA, 2002.

- [55] L. Borges, V. Marques, J. Bernardino, “Comparison of Data Mining techniques and tools for data classification”. In *C3S2E-13*, ACM, Porto, Portugal, 2013.
- [56] Y. N. Silva, S. W. Dietrich, J. M. Reed, “Integrating Big Data into the Computing Curricula”. In *SIGCSE’14*, ACM, Atlanta, GA, USA, 2014.
- [57] E. Trandafil, A. Allkoci, Elinda Kajo, A. Xhuvani, “Discovery and Evaluation of Student’s Profiles with Machine Learning”. In *BCI’12*, ACM 978-1-4503-1240-0/12/09, Novi Sad, Serbia, 2012.
- [58] B. Edwards, M. Zatorsky, R. Nayak, “Clustering and Classification of Maintenance Logs using Text Data Mining”. In *Proc. 7th Australasian Data Mining Conference (AusDM’08)*, Glenelg, South Australia, 2008.
- [59] L. Merschmann, A. Plastino, “A Bayesian Approach for Protein Classification”. In *SAC’06*, ACM 1-59593-108-2/06/0004, Dijon, France, 2006.
- [60] A. Veloso, W. Meira Jr., M. Cristo, M. Goncalves, M. Zaki, “Multi-Evidence, Multi-Criteria, Lazy Associative Document Classification”. In *CIKM’06*, ACM 1-59593-433-2/06/0011, Virginia, USA, 2006.
- [61] R. Frank, M. Ester, A. Knobbe, “A Multi-Relational Approach to Spatial Classification”. In *KDD’09*, 978-1-60558-495-9/09/06, Paris, France, 2009.
- [62] M. Ericsson, A. Wingkvist, “Mining Job Ads to Find What Skills are Sought After from an Employers’ Perspective on IT Graduates”. In *ITiCSE’14*, ACM 978-1-4503-2833-3/14/06, Uppsala, Sweden, 2014.
- [63] Q. Ding, Q. Ding, W. Perrizo, “Decision Tree Classification of Spatial Data Streams Using Peano Count Trees”. In *SAC 2002*, ACM 1-58113-445-2/02/03, Madrid Spain, 2002.
- [64] Ch. C. Aggarwal, “The Setwise Stream Classification Problem”. In *KDD’14*, ACM 978-1-4503-2956-9/14/08, New York, NY, USA, 2014.
- [65] Ch. C. Aggarwal, J. Han, Ph. S. Yu, “On Demand Classification of Data Streams”. In *KDD’04*, ACM 1-58113-888-1/04/0008, Seattle, Washington, USA, 2004.

- [66] Ch. C. Aggarwal, “Towards Exploratory Test Instance Specific Algorithms for High Dimensional Classification”. In *KDD’05, ACM 1-59593-135-X/05/0008*, Chicago, Illinois, USA, 2005.
- [67] J. Li, R. Topor, H. Shen, “Construct robust rule sets for classification”. In *SIGKDD’02, ACM 1-58813-567-X/02/0007*, Alberta Canada, 2002.
- [68] N. Jin, C. Young, W. Wang, “GAIA: Graph Classification Using Evolutionary Computation”. In *SIGMOD ’10, ACM 978-1-4503-0032-2/10/06*, Indiana, USA, 2010.
- [69] H. Fei, J. Huan, “Structure Feature Selection for Graph Classification”. In *CIKM ’08, ACM 978-1-59593-991-3/08/10*, California, USA, 2008.
- [70] A. Wegmann, “Theory and Practice behind the Course Designing Enterprisewide IT Systems”. In *IEEE Transactions on Education*, Vol. 47, No. 4, November 2004.
- [71] J. E. Froyd, Ph, C. Wankat, K. A. Smith, “Five Major Shifts in 100 Years of Engineering Education”. In *Proceedings of the IEEE*, 0018-9219, Vol. 100, May 13th, 2012.
- [72] Z. Shiller, “A Bottom-Up Approach to Teaching Robotics and Mechatronics to Mechanical Engineers”. In *IEEE Transactions on Education*, Vol. 56, No. 1, February 2013.
- [73] M. Marques, M. C. Viegas, M. C. Costa – Lobo, A. V. Fidalgo, G. R. Alves, J. S. Rocha, I. Gustavsson. In *IEEE Transactions on Education*, Vol. 57, No. 3, August 2014.
- [74] A. K. Kakar, “Teaching Theories underlying Agile Methods in a Systems Development Course”. In *47th Hawaii International Conference on System Science, IEEE*, 978-1-4799-2504-9/14, 2014.
- [75] M. G. Soto, R. Dzwonczyk, “Maximizing Service and Learning in an International Engineering Service Learning Program”. In *IEEE 2015 Global Humanitarian Technology Conference*, 978-1-4673-6561-1/15, 2015.
- [76] W. He, J. T. Kwok, J. Zhu, Y. Liu, “A Note on the Unification of Adaptive Online Learning”. In *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 28, No. 5, May 2017.

- [77] D. Klosters, “Matching Skills and Labor Market Needs”. In *World Economic Forum*, January 2014.
- [78] A. Ghani Kanesan Bin Abdullah, “Bridging the Gap between Industry and Higher Education Demands on Electronic Graduates’ Competencies”. In *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, Volume 8, Issue 1, 2013.