Machine Learning to Guide STEM Learning: Relative Importance of Social vs. Technical Competencies for STEM Students from Underrepresented Group

Karin Maurer, Technical University of Nuremberg, Germany
Heidi Schuhbauer, Technical University of Nuremberg, Germany
Patricia Brockmann, Technical University of Nuremberg, Germany

The IAFOR International Conference on Education – Hawaii 2020
Official Conference Proceedings

Abstract
Because STEM fields evolve quite rapidly, students interested in STEM professions are often unsure about what will be required of them upon graduation. In addition to specialized skills, such as programming, employers increasingly demand soft-skills, such as communication, self-reflection, conflict management and teamwork. Underrepresented groups in STEM include women, ethnic minorities and students from non-academic families. Academic planning can be more difficult, because they first generation students often lack role models as advisors. On the other hand, as minorities learning to integrate into the majority group, they may have learned to switch roles and see alternative perspectives, which represents a unique advantage to future employers. An analysis of university job market portals could help in career planning. Text mining of job ads enables the extraction of competencies necessary for entry-level positions. A job market portal supported by 15 German universities is analyzed using statistical and machine learning tools: linear regression and neural networks. The analysis of over 22,000 job ads over 14 years enables the identification of specific competencies desired by potential employers in STEM fields. By tracking changes in employer demands, trends can be identified of which skills have become more and less desirable over time. This analysis found that social competencies have become more important. The probable future importance of individual qualifications can be forecast, to help students from underrepresented groups in their academic planning. This work is part of a larger research project to recruit, retain and support STEM students from underrepresented groups.

Keywords: STEM, diversity, soft skills, technical, competencies, machine learning

Introduction

The digital revolution has permeated all aspects of our modern society. Both low-skilled manufacturing jobs as well as mid-level administrative employees are increasingly being replaced by automation, robotics and artificial intelligence (Bessen 2016). At the same time, industrialized nations are suffering from a shift in demographics due to low birth rates. The number of new graduates in STEM subjects (Science, Technology, Engineering and Math) has not kept up with the increasing demand for qualified professionals. The skills gap between under-qualified job seekers and the increasing number of unfilled positions for highly qualified STEM workers continues to increase (Doerschuk, 2016).

Historically, people from certain demographic groups have been quantitatively under-represented among university graduates in STEM subjects: first generation students, people who come from a migration background, single parents and women.

Instead of viewing digitalization solely as a threat, this research focuses on the advantages which artificial intelligence could bring. This work addresses two research questions:

1. What are the important specialized skills and social competencies graduates in STEM majors need to master for their future careers in a digitalized society?
2. Can machine learning methods be useful in analyzing large quantities of unstructured data about the job market to derive information about which professional competencies will be needed in the near future?

Section 2 Related Work discusses some of the literature about career competencies in STEM subjects. Section 3 Methodology describes the machine learning methods employed in this investigation. Section 4 Results presents the outcomes of these experiments. Section 5 Conclusions discusses the implications of these results for STEM students. Finally, Section 6 presents plans for future research.

Related Work

This section provides an overview of some of the literature related to the current proportion of STEM students from underrepresented groups, their career goals and the competencies which will be required of future graduates in STEM subjects. The complex interplay of language, cultural and gender diversity are also discussed.

First generation students from non-academic households can often be subject to different socialization and cultural norms than students whose parents both have university degrees. Pupils from academic households grow up with access to economic capital, experience geographic mobility and are often socialized to emphasize independence, personal choice and self-expression (Miller 2005). First generation students, in contrast, are often raised to respect cultural norms which encourage interdependence, placing the needs of others before one's own and maintaining strong family bonds. Working-class families frequently suffer from limited economic capital, environmental constraints, with few opportunities for choice, control and influence (Grossmann 2011). Stephens, et al., postulate a theory of "cultural mismatch" (Stephens 2012), which describes cultural obstacles which first generation students encounter as they begin their university studies. Students from underrepresented groups often perceive themselves as “not belonging”. Lack of identification can be a major barrier to underrepresented students participating in STEM subjects (Sinclair 2014). Cross (2001) observed students in a highly competitive engineering degree program, who had not been socialized to act independently and to be self-reliant. These students initially had problems with self-esteem and persistence. These problems could be partially mitigated by high levels of social support among their peers.

A large number of authors have analyzed the effect of gender on attitudes toward STEM subjects in schools and in the workplace. Environmental and cultural influences as well as stereotypical views of STEM subjects can prevent women from entering STEM fields (Blum 2007). According to information from the international PISA study in 2015 (OECD 2016), girls and boys performed similarly on tests of general science literacy in most nations. Although this would imply that boys and girls possess equal innate abilities in STEM subjects, women obtained fewer university degrees than men in STEM disciplines (WEF 2015).

Stoet and Geary found that countries with a high level of gender equality tend to have some of the lowest proportions of women in STEM subjects. Gender equality was measured with the Global Gender Gap Index (WEF, 2015). They call this phenomenon the “education-gender- equality paradox” (Stoet, Geary 2018). They deduced that countries with the highest degree of gender equality tend to be welfare states, with a high level of social security for all of their citizens. Countries with lower gender equality tend to have less economic security and more difficult living conditions. They postulate that individuals in countries with low gender equality would place more value on high paying STEM occupations, which could provide them with economic security. Their analysis suggests that in more gender equal welfare states, the potential financial costs of foregoing a STEM career amplify intra-individual academic strengths. Faulkner (Faulkner 2009) discusses the subtle dynamics which can contribute to a feeling of “belonging” in work relationships. She discusses the importance of informal conversation topics among colleagues, which can make women and other underrepresented groups feel like outsiders.

The intercultural competencies necessary to work in international engineering teams have been investigated by a number of authors. Beecham, et al., (Beecham 2017) identified a number of challenges inherent to working in international development teams, which new STEM graduates will need to address: distance, teamwork, soft issues, stakeholders, infrastructure and distributed development processes. They categorized various types of distances, such as physical (geographic), time zones, cultural, language and institutional distances. Other authors, such as Hoda et al., (Hoda 2016) concentrated on the socio-cultural capabilities which students need to learn to work effectively in globally distributed teams. They pointed out the importance of overcoming language barriers, different perspectives regarding time, attitudes towards achievement, differences in autonomy and work habits as well as assumptions about national culture. They underline the importance of cross-cultural training. One example of the importance of cultural sensitivity in requirements engineering was reported by Hinze, et al. (Hinze 2018). To develop a medical app aimed at improving the health of migrant communities, sensitive medical data needed to be collected. When dealing with multicultural stakeholders, it is of utmost importance to first establish a sense of trust. Ideally, they recommend that one member of the research team should come from the cultural community studied. This bi-cultural individual can help build to bridges between diverse cultural expectations.

This overview of research demonstrates some of the challenges which students from underrepresented groups will face upon graduation. The question arises as to whether these non-traditional students can leverage their experiences of belonging to minority groups in order to make unique contributions to increase the diversity of perspectives examined when solving innovative problems. Ilumoka (Ilumoka 2012) discusses the importance of diversity in engineering teams. Especially during the requirements engineering phase, nontechnical skills, such as intercultural communication and foreign language abilities, can be of exceptional value for multi-national teams or for stakeholders in foreign countries. During the development phase, cooperation, team-building and conflict management skills can prove vital for the success of an engineering project.

Methods

In order to answer the research questions about which technical and social competencies are most important and to examine whether machine learning methods can be useful in analyzing large amounts of unstructured data from the labor market, it is first crucial to define a pool of test data and then to choose the appropriate analysis methods.

The first step is to define a data pool which contains sufficient information about the labor market and from which it can be deduced which vocational competencies will be needed in the near future. The higher the relevance and the quality of the data, and the better the data preparation before analysis, the more useful the subsequent results of the analysis will be (Mohri, Rostamizadeh, & Talwalkar, 2012, p. 1). Particularly for the first research question, concrete information on competencies is needed. One source of information on competencies required for a particular job can be found in job advertisements. For this reason, the job advertisements of the university career portal database, which contains job advertisements from 2003 onwards, which are used for this research. Since the university job exchange also contains job advertisements for non-STEM occupations, a subset of the data was pre-selected, so that only job advertisements for STEM occupations were considered.

After selecting the appropriate data pool, the next step is to pre-process the data. To be able to perform analyses with a large amount of unstructured data, the data must be available in a certain form. In most cases, the data is never available in a format which can be analyzed directly. Since job advertisements are usually formulated as running text, and required competencies are usually found in unstructured enumerations, the job advertisements must first be subjected to pre-processing. This ensures that the data is available in such a way that a computer algorithm can analyze it. The pre-processing of the data is a lengthy process in and of itself. Thus, from the selection of the data up to the evaluation of the results, it usually requires between 60 and 70 % of the total time expenditure (Maurer, 2019, p. 108).

In order to answer the question about which technical and social competences are most important, a list of all the competences in question from the job advertisements must be available, as well as an indication of how often these competencies were mentioned in total. This allows for an assessment of how important a single competency is across the multitude of job advertisements.

The second research question on the competencies needed in the near future requires a slightly different data structure. Since this requires a prognosis, time series have to be used. It has to be calculated how often a competence occurred in each year in the past.

The processing of the data starts with tokenizing. In tokenizing, the words and punctuation marks are separated from each other, so that each word and each character can be considered independently. All characters and numbers that are not alphanumeric and the so-called “stop words” (e.g. filler words such as "der", "die", "das") are then removed from these words. Some of the remaining words can then be further ruled out. In the case of job advertisements, for example, company names can be removed from the advertisements, because they are not relevant for the identification of requirements.

Parallel to the pre-processing of the data, “dictionaries” must also be created, which contain as complete a scope of all of the competencies as possible. The competencies from the dictionaries are then individually combined with the words from the job advertisements. If it is recognized that a word from the job advertisements is equal to a word from the dictionary, it is marked. This, however, carries the risk of redundancies, e.g. due to different conjugations used for a word in the job advertisements. In order to avoid redundancies, both the words in the job advertisements and the words in the dictionaries are stemmed. Stemming means reducing a word to its root. After stemming the words, the actual information extraction can take place, on the basis of which the analysis can then take place.

In order to obtain as good and usable analysis results as possible, a lot of historical data is required. (Mohri et al., 2012, p. 1). It is also necessary to use the same number of data records every year. Since this work is based on the number of occurrences of a competency, a different number of data sets per year would falsify the result. The years 2003 and 2004 from the career services database, for example, have less than 100 data sets, while some years from 2005 onwards include approximately 10.000 data sets. For this reason, only the years 2005 to 2018 were used in this analysis.

	year	occurrence	competence	type of competence
0	2018	1114	Erfahrung	0
1	2017	1029	Erfahrung	0
2	2013	967	Erfahrung	0
3	2015	956	Erfahrung	0
4	2014	956	Erfahrung	0
1791	2011	1	Adobe	0
1792	2009	1	Adobe	0
1793	2007	1	Adobe	0
1794	2013	1	umweltbewusst	1
1795	2017	1	tolerant	2

Figure 1: Results of the data processing

Figure 1 shows an example of the data used to identify competencies needed in the near future. It is a list of all competencies included in the job advertisements. This list is enriched by information on how often they are mentioned and in what year. The example in Figure 1 shows that the competency "Erfahrung" (experience) was requested in the years 2018, 2017, 2013, 2015 and 2014, with the corresponding frequencies listed in the first column. In order to show that other competences were also mentioned in the job advertisements, lines 1791 to 1795 show further entries in the list. An aggregation of the list per competency enables the determination of which technical and social competencies were important for STEM majors.

However, the presentation of the data in Figure 1 is not yet suitable for answering research question two, namely whether machine learning methods can be useful to analyze large amounts of unstructured data from the labor market. Mitchell describes machine learning as suitable for analyzing large amounts of data and generating knowledge from the data that is not yet available (Mitchell, 2010, p. 14). But it is not enough to know that new knowledge can be generated. It is also important to be able to assess whether the models can deliver meaningful results. Therefore, two different models will be tested. The results of both analyses should then be compared with each other in order to be able to make a statement about the functionality of the models. The machine-learning models are to provide forecasts for the years 2017 and 2018 on a trial basis, so that these can be compared with the actual figures from the list in Figure 1.

Each model must first be fed with training data for analysis, so that the algorithm can learn from it. This procedure is called "training," because the machine learning models become better and better through the training data by adapting themselves to the data available (Mitchell, 2010, p. 17). This is how new knowledge can be generated from current data (Müller & Guido, 2017, p. 123). After training the respective model, the forecast figures for the years 2017 and 2018 are to be provided. This data, which is used to check the model's functionality, is called test data. A comparison with the actual figures from 2017 and 2018 now makes it possible to assess the functionality of the models.

When training a machine learning model, it is important to use as much training data as possible, because the more training data you use, the more accurate the prediction result will be. Therefore, the years 2005 to 2016 are the training data and the years 2017 and 2018 are the test data. The model that provides the better forecast data is then used for the actual forecast that provides information about skills needed in the near future.

Now that the procedure for evaluating the models has been determined, two machine learning models are next selected for comparison. The first model selected is linear regression, a classical machine learning model. Linear regression is a machine learning model that mathematically investigates the relationship between two characteristics (Teschl & Teschl, 2007, p. 215) and uses this mathematical correlation to predict future values. One characteristic is dependent on the other (Raschka, Mirjalili, & Lorenzen, 2018, p. 317). In the case of job advertisements, the number of requests depends on the year in which they were made. Linear regression is a method of supervised learning which calculates continuous values using a regression line (Raschka et al., 2018, p. 317). The future values can be read off from the regression line. Since linear regression is a model that provides continuous values, the years 2005 to 2016 can be used as training data without further adjustments. The years 2017 and 2018 can be used as test data. The regression line resulting from the training and test data can be used to read the forecast data and compare it with the test data.

The second method selected for evaluation is a neural network, which calculates future values on the basis of an algorithm. The recognition of correlations requires the use of multi-layered neural networks. A neural network consists of neurons that work in parallel and send information to each other via directed connections (Kruse et al., 2015, p. 13). The algorithm is fed with input values and the subsequent output values are calculated by layers and different weightings of the neurons. In order to generate meaningful results, neural networks must be trained. This means that the neural network is fed with input values. The network then back-propogates during the execution of the algorithm to check whether the results are valid. If these are not valid, the algorithm adapts to provide better results. Once the network is trained, it can be used to make the actual predictions. Since the neural network learns iteratively and uses data from previous years, the net must be trained somewhat differently than a linear regression model. More specifically, this means that for the neural network the years 2005 to 2014, and parallel to this, the years 2006 to 2015, each one year later, must be used as input values. The years 2016 and 2017, and 2017 and 2018 respectively, then serve to verify the forecast values by the neural network.

Figure 2 below illustrates the distribution of test and training data for the neural network. It can be seen that both the training data and the test data are offset by one year so that the algorithm can learn.

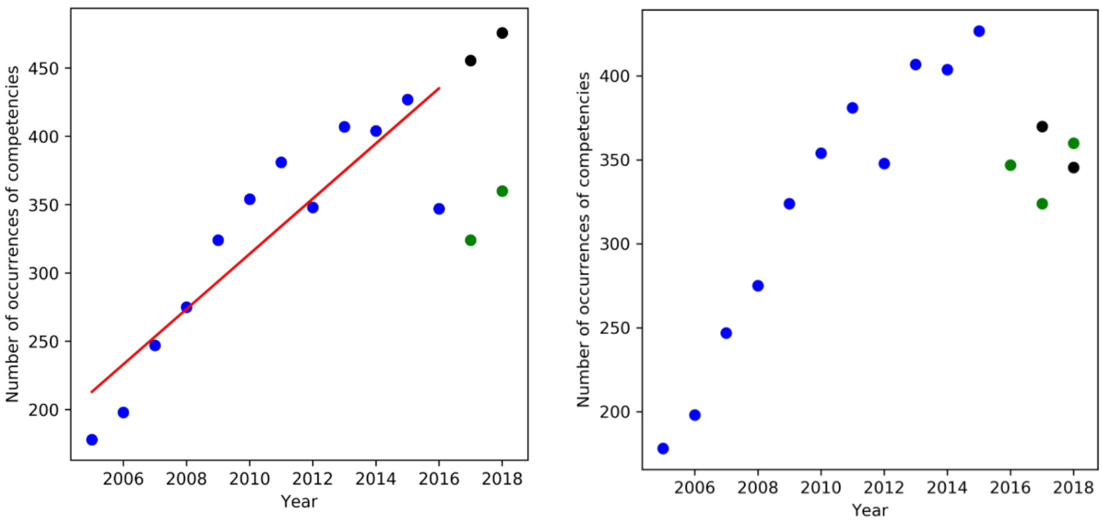
year	training data X	training data y	test data x	test data y
2005				
2006				
2007				
2008				
2009				
2010				
2011				
2012				
2013				
2014				
2015				
2016				
2017				
2018				

Figure 2: Partitioning of the training and test data for the neural network

After both the regression model and the neural network models have been executed, the forecast values for 2017 and 2018 are compared with the actual available values. The selection is made on the basis of the percentage difference between the two values, so that the better machine learning model is used for the actual forecast for the near future.

Results

Figure 3 shows an example of the result of the first analysis for selecting a machine learning model. The left side of Figure 3 shows the result of the linear regression, the right side shows the result of the neural network. Both diagrams are based on the performance of the competence analysis "independently". The green dots are the actual figures, the black dots the forecast figures for the years 2017 and 2018. Figure 3 shows that for the competence "independent" the neural network worked better than the linear regression. A total of 12 analyses per machine learning model were carried out to select the appropriate model. The evaluation of the 24 analysis results per model (12 each for the years 2017 and 2018) showed that the neural network delivered better results in 18 of 24 cases, which is why the neural network was used to forecast future competences.



Linear Regression Neural Network

Figure 3:Example of the prediction for the competence “independence”

First of all, when presenting the results, one should answer the research question:

1. What are the important specialized skills and social competencies that STEM graduates need for their future careers in a digital society?

Specialized Skills		**Social Competencies**	
Experience	12,200	Ability to work in teams	5,214
University Degree	9,199	Self-reliance	4,574
Excel	3,731	Analytical abilities	3,060
English	2,781	Commitment	2,882

Table 1: Most important specialized and social competencies

Table 1 shows the four most important specialized skills and social competencies that STEM graduates should master in their future careers. The most commonly requested specialized skill is the experience that graduates should bring with them. However, this must be viewed with caution. Especially graduates from underrepresented groups often do not have time to work part-time to gain work experience during their studies. In particular, students from non-academic families have to deal intensively with everyday student life and cannot expect the same support from home as students from academic families. Gaining work experience is of secondary importance to their studies. As this is an analysis for graduates from STEM courses, all students can demonstrate this technical competence. Specialized skills in Excel comes in a distant third. The mastery of Excel cannot be completely guaranteed by STEM majors. However, Excel is often used as a tool, so at least the basics are taught. The fourth specialized skill identified as one of the most important is English as a second language. For STEM graduates at the TH Nuremberg in Germany, this is a skill which should be readily mastered, as there are two English lectures in the bachelor's degree program and even a third in the master's degree program.

The most important social skills are led by the ability to work in a team, followed by self-reliance. These two competencies are promoted by learning methods in which group work is often carried out during courses and a lot of independence is expected in order to obtain the academic title. The third most important social competency is analytical ability. This competence is promoted in courses such as "Software Engineering", where, for example, customer requirements have to be analyzed and structured. The final important competency listed for a future career in a digital society is commitment. However, this does not say anything about the extent to which these competencies are pronounced among the students. This aspect will be further discussed in the conclusions.

After the identification of the most important specialized skills and social competences, the second research question can now be answered:

2. Can machine learning methods be useful to analyze large amounts of unstructured data from the labor market in order to derive information on which vocational competencies will be needed in the near future?

The answer to this research question was initially preceded by general analyses.

Figure 4 below shows a diagram of the number of occurrences of competencies, broken down by year. The graph shows that the number of different competencies has increased by about 20 since 2005. Since 2007, the number of different competencies has fluctuated between approx. 125 and approx. 137. In principle, the diagram in Figure 4 shows that the required range of different competencies has increased since 2005 and has fluctuated since then. However, until 2018, the number of different competencies required has not exceeded 135.

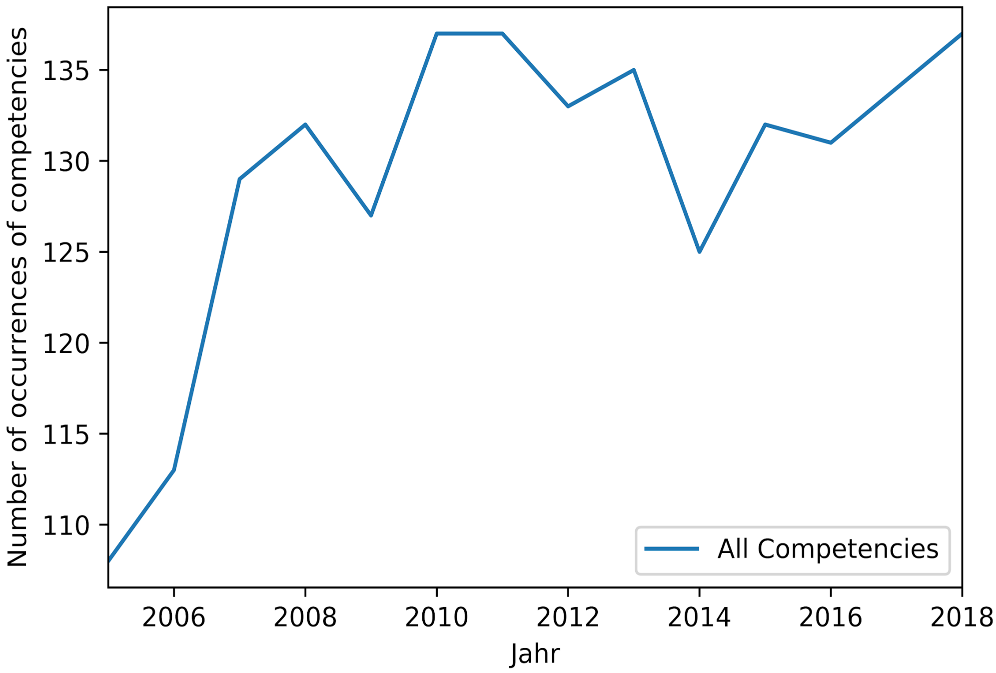


Figure 4: Analysis of the number of competencies pro year (Maurer, 2019, p. 62)

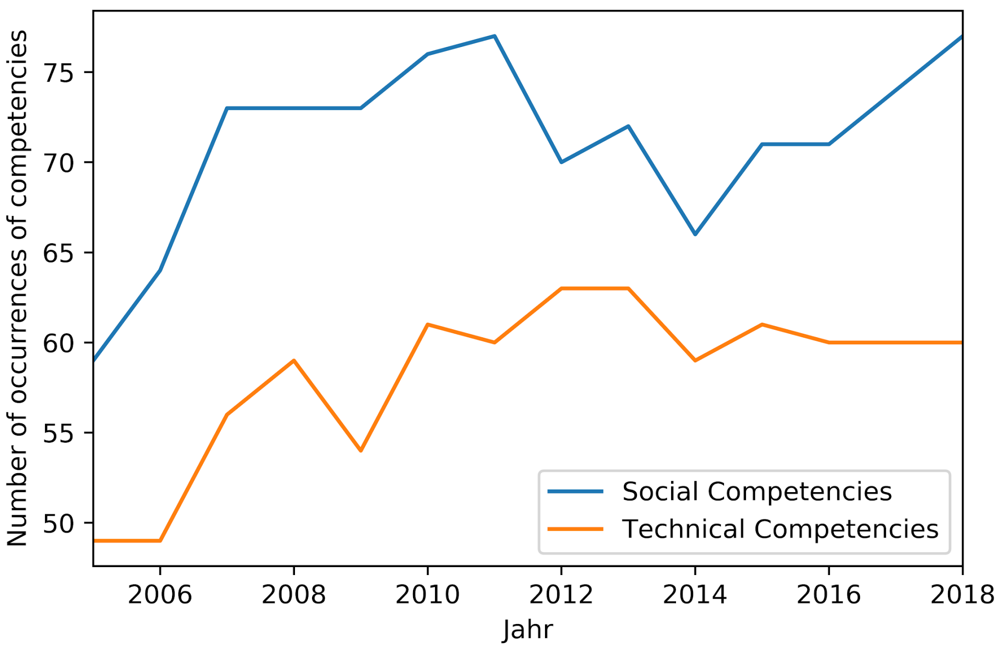


Figure 5: Comparison of social competencies vs. specialized skills (Maurer, 2019, p. 64)

In addition to the total number of competencies in Figure 4, Figure 5 shows a comparison of the number of technical and social competencies. The upper line shows the number of social competencies, the lower line the number of technical competencies. It can be seen that already at the beginning of the data series in 2005, the required number of social competencies is higher than that of technical competencies. This continues until 2018. In addition, it can be seen that the demand for technical competencies has apparently stagnated since 2016. In contrast, the demand social competencies has been rising steadily since 2016. At the same time, the trend in the variation of social skills required continues to rise.

Looking at the diagrams in Figures 4 and 5, it can therefore be said that social skills have become increasingly more important in the occupational environment than technical skills. This is underlined in particular by the available data, since it concerns data on STEM occupations, i.e. exclusively technical and scientific occupations. The assumption made by society that only technical know-how counts in technical occupations was refuted by Figure 5 above. In particular, this also has an advantage for underrepresented groups. Technical skills can be learned, while soft skills are often difficult to acquire. Since students from underrepresented groups have often had to master a large number of soft skills in the course of their lives, e.g. for integration into majority groups, there is a potential advantage here. The know-how for technical competencies can then be acquired during an intensive learning process.

	Experience	**University Degree**	**Excel**	**English**
Occurrences 2018	1114	801	207	170
Forecast 2020	1201	838	195	164
	Teamwork	**Communication**	**Networking**	**Cooperation**
Occurrences 2018	359	105	14	5
Forecast 2020	358	101	11	5
	Self-Reliance	**Analytical Abilities**	**Commitment**	**Proactive**
Occurrences 2018	360	207	177	212
Forecast 2020	379	205	173	219

Table 2: Forecast for 2020 in comparison to actual occurrences in 2018

In addition to the general analyses, the forecast for the near future for the required vocational competencies will now be presented. The machine learning algorithms were trained in such a way that they provided forecasts for the year 2020. Table 2 presents the concrete forecast figures for 2020, compared to the figures already available from 2018 and shows the result of the forecasts for the year 2020. Since the machine learning algorithms used here can only be executed for one competency at a time, 12 competencies were selected for which a future forecast should be prepared (Maurer, 2019, p. 66). Table 2 above shows that two of the subject-specific competencies, namely "experience" and "degree", are becoming more important, while the "Excel" and "English" competencies are less required. In the area of social competencies, the forecast shows that expectations will not change for the competency "cooperation", but that demand for the competencies "ability to work in a team," "communication" and "networking" will decrease. The results for personal competencies are similar to those for technical competencies. The two competencies "self-reliance" and "proactive" are forecast to experience an increase in demand, while the competencies "analytical ability" and "commitment" should be slightly less in demand. In summary, the concrete forecasts for the 12 competencies show that two technical competencies are becoming more important and two less important. On the other hand, five of the eight soft skills will be less in demand, while one remains the same and two will be more in demand. The figures from Table 2 contradict the tendencies from the diagrams in Figure 4 and Figure 5. However, it should be noted that only 12 of the 135 different competencies identified were considered and thus no reliable result can delivered based on Table 2.

Competence	Forecast
Experience	1,201
University degree	838
Self-reliance	379
Teamwork	358
Proactive	219
Analytical ability	205
Excel	195
Commitment	173
English	164
Communication	101
Networking	11
Cooperation	5

Table 3: Necessary competencies in the near future

Table 3 above provides the concrete answer to research question two. Specifically, experience and study will continue to be required. However, independence and the ability to work in a team will also be useful in the future. In addition, personal initiative is welcome, followed closely by analytical skills. However, classical skills in Excel and commitment will continue to be important. As globalization continues to progress, English skills and communication skills are also important. Good networking will also be required. Finally, graduates seeking their first entry-level jobs should show a high level of cooperation.

Conclusions and Future Work

In conclusion, this study has been able to answer two research questions:

1. What are the important technical and social competencies graduates in STEM majors need to master for their future careers in a digitalized society?

The most widely sought technical skills were: work experience, a university degree, proficiency in Excel and proficiency in English as a second language. The most widely sought soft skills were: ability to work in teams, self-reliance, analytical abilities and commitment.

2. Can machine learning methods be useful in analyzing large quantities of unstructured data about the job market to derive information about which professional competencies will be needed in the near future?

Machine learning methods have shown to be useful in analyzing large amounts of unstructured data on the labor market. The machine learning methods forecast the following professional competencies as important in the near future: work experience, a university degree, self-reliance, ability to work in teams, proactive, analytical ability.

These results leads to a future research question: To what extent can students from under-represented groups fulfill these competencies? Underrepresented groups may have an advantage in acquiring these soft skills. Especially students from underrepresented groups often have to develop self-reliance, as they may study at a university far away from home and may therefore be more independent than the average student. In order to integrate into a culturally foreign environment, not only a great deal of communication talent and commitment, but also personal initiative is required. This could enable first generation students, students with a migration background, single parents or women in STEM to develop higher levels of social competency.

This work is part of a larger research project, named “DiaMINT”. The goal of the project is to recruit, support and retain students from underrepresented groups in STEM subjects. DiaMINT covers each phase of the student customer journey, from the initial information gathering phase, through application, admission, orientation, internships, exams, theses and entry into the job market (Schuhbauer 2019).

Acknowledgements

This work was supported by a research grant for the project DiaMINT by the Staedtler Trust.

References

Beecham, S., Clear, T., Barr, J., Daniels, M., Oudshoorn, M. and Noll, J., 2017, Preparing Tomorrow's Software Engineers for Work in a Global Environment, IEEE Software, Number 1, Vol. 34, pp. 9-12.

Bessen, James E., How Computer Automation Affects Occupations: Technology, Jobs, and Skills (October 3, 2016). Boston Univ. School of Law, Law and Economics Research Paper No. 15-49.

Blum, L., Frieze, C., Hazzan, O., Dias, M. B. , “A Cultural Perspective on Gender Diversity in Computing”. In: Reconfiguring the Firewall, P. Meszaros, A.K. Peters Eds., CRC Press, 2007, (S. 109-133).

Cameron, D. (2001). Working with spoken discourse. London: SAGE.

Coates, J. (2007). Talk in a play frame: More on laughter and intimacy. Journal of Pragmatics, 39, 29-49.

Cross, S. E., & Vick, N. V. (2001). The Interdependent Self-Construal and Social Support: The Case of Persistence in Engineering. Personality and Social Psychology Bulletin, 27(7), 820–832. https://doi.org/10.1177/0146167201277005

Doerschuk, P., Bahrim, C., Daniel, J. et al., (2016). Journal of Science and Education Technology, Vol. 25, No. 4, pp. 682-695. https://doi.org/10.1007/s10956-016-9622-8

Drew P., & Heritage J. (1992). Analysing talk at work: An introduction. In P. Drew, & J. Heritage (Eds.), Talk at work (pp. 3-65). Cambridge: Cambridge University Press.

Faulkner, W., (2009). Doing gender in engineering workplace cultures. I. Observations from the field. Engineering Studies, 1:1, 3-18, DOI: 10.1080/19378620902721322

Grossmann, I., & Varnum, M. E. W. (2011). Culture, social class, and cognition. Social Psychological & Personality Science, 2, 81–89. doi:10.1177/1948550610377119

Hinze, A., Timpany, C., Bowen, J., Chang, C., Starkey, N., and Elder, H., 2018, Collecting Sensitive Personal Data in a Multi-Cultural Environment, HCI 2018 Proceedings of the International Human Computer Interaction Conference, Las Vegas, Nevada, U.S.A.

Hoda, R., Ali Babar, M., Shastri, Y., Yaqoob, H., 2016, Socio-Cultural Challenges in Global Software Engineering Education, IEEE Transactions on Education, No. 3, Vol. 60, pp. 173-182.

Ilumoka, A. 2012, Strategies for overcoming barriers to women and minorities in STEM, IEEE 2nd Integrated STEM Education Conference (ISEC), Ewing, New Jersey, U.S.A., pp. 1-4.

Kruse, R., Borgelt, C., Braune, C., Klawonn, F., Moewes, C., & Steinbrecher, M. (2015). Computational Intelligence: Eine methodische Einführung in künstliche neuronale Netze, evolutionäre Algorithmen, Fuzzy-Systeme und Bayes-Netze (2., überarbeitete und erweiterte Auflage). Computational Intelligence. Wiesbaden: Springer Vieweg. Retrieved from http://dx.doi.org/10.1007/978-3-658-10904-2 https://doi.org/10.1007/978-3-658-10904-2

Maurer, K. (2019). Kompetenzentwicklung in MINT-Berufen: Ermittlung zukünftig notwendiger Kompetenzen mittels Machine Learning (Masterarbeit). Technische Hochschule Georg Simon Ohm, Nürnberg.

Miller, P. J., Cho, G. E., & Bracey, J. R. (2005). Working-class children's experience through the prism of personal storytelling. Human Development, 48, 115–135. doi:10.1159/000085515

Mitchell, T. M. (2010). Machine learning (International ed., [Reprint.]). McGraw-Hill series in computer science. New York, NY: McGraw-Hill.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). Foundations of machine learning. Adaptive computation and machine learning. Cambridge, Mass., London: The MIT Press.

Müller, A. C., & Guido, S. (2017). Einführung in Machine Learning mit Python: Praxiswissen Data Science (K. Rother, Trans.) (1. Auflage). Heidelberg: O'Reilly. Retrieved from http://proquestcombo.safaribooksonline.com/9781492064619

OECD (2016). PISA 2015 assessment and analytical framework: Science, reading mathematic and financial literacy, Paris, France.

Raschka, S., Mirjalili, V., & Lorenzen, K. (2018). Machine Learning mit Python und Scikit-learn und TensorFlow: Das umfassende Praxis-Handbuch für Data Science, Deep Learning und Predictive Analytics (2., aktualisierte und erweiterte Auflage). Frechen: mitp Verlags GmbH & Co. KG.

Schuhbauer, H., Brockmann, P. (2019). "Digitization of the Student Life Cycle to promote under-represented Groups in STEM Subjects", 16th International Conference Cognition and exploratory Learning in the Digital Age, CELDA 2019, Cagliari, pp. 249-256.

Stephens, N., Fryberg, S., Markus, H., Johnson, C, Covarrubias, R., Unseen Disadvantage: How American Universities' Focus on Independence Undermines the Academic Performance of First-Generation College Students, Journal of Personality and Social Psychology Vol. 102, No. 6, 1178-1197, 2012.

Stoet, G., & Geary, D. C. (2018). The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. Psychological Science, 29(4), 581–593.

Teschl, G., & Teschl, S. (2007). Mathematik für Informatiker: Band 2: Analysis und Statistik (2. Auflage). eXamen.press. Berlin, Heidelberg: Springer-Verlag Berlin

Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-540-72452-0
https://doi.org/10.1007/978-3-540-72452-0

World Economics Forum WEF (2015). The global gender gap report 2015. Geneva, Switzerland.

Contact email: patricia.brockmann@th-nuernberg.de