Using NLP to bridge Data Science Skills Gap in Namibia. A Survey

Tokolo N Kashupi
Department of Informatics
Namibia University of Science and Technology
Windhoek, Namibia
ndiyaholek@gmail.com

Edmore Chikohora
Department of Informatics
Namibia University of Science & Technology
Windhoek, Namibia
echikohora@nust.na

2021 3rd International Multidisciplinary Information Technology and Engineering Conference (IMITEC) | 978-1-6654-1749-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/IMITEC52926.2021.9714682

Abstract:
Previous studies confirm that many graduates are struggling to get jobs in the industry as they lack the required skills. Finishing high school and going to university was supposed to be the way of opening up job opportunities, which is no longer the case. The gap between industries' required skills and Higher Education Training methods is huge, even after past studies tried to find out where the problem lies. In this study, a literature survey was used to assist with the identification of existing shortcomings in training institutions, in particular the misalignment of Higher Education Training with the Industry expectations. To help understand Natural Language Processing (NLP) and how it has been applied before to solve human challenges problems. The ultimate goal is to design a model that helps bridge the skills gap between academic institutions and the industry. This model will have a component of mining data from existing curriculum documentation of Data Science, the results are to be reviewed and compared to the industry expected skillset to help design a preferred curriculum. The study used methods of research, analysis and data collection based on the mixed method of qualitative, survey distribution for deep understanding on how they are picking up this skills gap and what they recommend should be done to bridge the gap. This study is mainly focused on the Computing faculty in the Namibia University of Science and technology and First National Bank Namibia, which are considered to have or plan to have a data science department in the years to come. The results from this research are significant as they align the academic system to the skills required in the industry.

Keywords— Data Science skills, higher education, Natural Language Processing, Data Science, industry needs, knowledge engineering, skills gap, graduates.

I. Introduction

Academic institutions' curriculums do not fit the industry's needs. Studies have been conducted and researchers have tried to find ways on how to improve traditional education by bridging the existing gap between the academic institutions' education system and the industry needed skillset. According to [1], academic institutions' programs are theoretical topics that are stated to be the root cause of the huge existing gap between academic institutions' education and the industry needs for employment.

"A Limited number of skilled professionals in the industry are not caused by individuals not wanting to call themselves scientists, engineers, and analysts, etc. but it is the shortage of one quality aspect of well-studied and well-experienced professionals in a specific field", says [1]. About 69 higher Education Training institutions that offer data science courses including those that develop data science curriculums were surveyed/interviewed about their curriculums, tools used, and challenges faced when teaching Data Science courses [1]. From a computer perspective, in 2020, Namibia University of Science and Technology designed their "Big Data Technologies" short course to look at topics like data storage, Visualizations, statistics methods etc. through high-performance methods. The study will incorporate Natural Language Tool kit in Python that helps users build tools with sophisticated linguistic processing [2], in this case, design the framework.

This study is aimed at surveying literature to examine previous studies on bridging the gap between higher education and the industry needs and give readers an idea of the current research study and review examining how previous researchers have dealt with bridging the industry-academia gap, the challenges of bridging the gap of higher education, the courses of the skills gap in the data science career and industrial skills in Namibia and globally. The rest of the paper is organized as follows; the background information about the study, survey results and discussions then the last chapter concludes the paper with recommendations on future work.

II. Background Study

Data is the world's most expensive asset which is turning organizations into analytics-intensive enterprises and will provide more job offers in the future. To manage the valuable asset of data, data scientists with relevant industry-specific skills will be the most needed experts worldwide. Data Science includes components of NLP, Machine learning, Business Intelligence, which provides business organizations with a range of enterprise technologies for visualizing data, deriving useful insights, and developing smart business solutions. [3] stated that industry organizations need "professionals who possess a great understanding, have current relevant analytical methods, be familiar with the technology, can interact, analyze, communicate and can present their work effectively" (p. 25).

Among the components of Big Data, scrapping data from online sites, extracting information from text or voice data seem to be still challenging to date. It is quite overwhelming when it comes to analyzing data to get information from either social media posts, news etc. In this case, getting data science skills from the web to present to HET for they are required from graduates [4]. This is where NLP tools come in, e.g., Natural Language Toolkit – a collection of language processing open-source tools in python that is recommended to be suitable for researchers, students, etc. [5]. NLP tools capture texts called (sentiment analysis), which can be unlimited and identify the keywords from that text called (tokenization/summarization). NLP can be referred to as the area of Artificial Intelligence in computer science that is known to give computers the ability to understand and respond to text and speeches like humans do [5].

Looking at bridging the DS skills gap between HET and the industry, NLP could be used for extracting online DS course contents and program curriculum from the internet; Extract data science and analytics industry jobs requirements from online sources. Using NLP tool kits, comparisons can be done between different data sets by finding key terms and phrases. With this approach, relationships between DS program curricular and job requirements can be established.

[2] developed an IBM Watson cognitive system, a natural language question-answering question, in an education context. "The system determines scores the answers based on the knowledge it has picked up. It uses Natural language to understand questions to generate and evaluate the hypotheses then produces the answers and supports them with evidence." [6] developed a machine that detects languages in online resources like videos and other voice recourses to the Indian language. Since the recognition and synthesis of text-speech, machine translation etc. have become the most popular technologies of Natural Language Processing and machine learning. There has been several NLP machines and programs developed and most popular with IBM Watson's Natural Language called "The NLP service for advanced text analytics [5].

The question then becomes, are the universities and the industry prepared for this? According to [7], the curriculums need to be designed in a way that caters for specific education needs in all fields of education, including Data Science and other new upcoming industry professions. Technology is evolving and it is the growing field of science that requires a strong and wide building approach that includes data-driven technologies like Data Science.

A. *Possible causes of the skills gap between HET and the Industry*

The skills gap between Higher Education Training and the working industry has been going on for at least 2 decades ago to date, which makes individuals and job seekers question as to what is the clear reason to the cause of this gap? [19]. Data Science education at the level of Higher Learning lacks the most common and accepted model of teaching that reflects the entire lifecycle of data handling, to grow and build data-driven scientists and an improved digital economy [6]. Data science professionals are becoming a demand in Namibia, although there is a lack of expertise and education. Most of it is caused by the lack of awareness of the data science field of expertise. In the US [14], PWC has done a study on what's missing and what needs to be invested in an American student. The current pool of job skills shortage will not satisfy the needs of the growing business usual strategies. Employers and educators need to transform to grow the potential and promise of producing well equipped and skilled data scientists by looking at eight different strategies that can equalize the demand and the required skills in the field of data science. This theory could be applied too in Namibia.

[8] looked at the causes of these skills gap using the question "Are faculty views on the undergraduate CS program the main cause of the skills gap?", and found out that improving employability is not just the common goal when designing the curriculum, which could be the case in Namibia. The curriculum is mostly developed without looking at the student's main interests in attending universities. They picked up three other reasons from the faculty views which are the causes of the industry-academia gap. One – "the faculty believes that, above the computer Science fundamentals, for example, the goals of the Computer Science degree need to be defined based on the students' objectives." This is why the curriculums are now set but not in favor of the students' goals and objectives. Two – "there are always debates during the curriculum design process of faculties either opposing or supporting the industry preparations of students upon degree completion". And finally three – "faculty formally agreed to have been struggling with proper preparations of students for the industry looking at the class size, lack of background industry practices among the faculty members. Having difficulties when it comes to designing real-time projects without overworking both students and instructors". With all the above-mentioned reasons, it makes much sense why the gap exists and Data science course is still not offered in the Namibian Universities. It also concludes that the Universities management are the main causes of the gap, mainly because they have fear of facing either the number of students enrolled or lack of industry experience. Another reason being, due to the lack of proper institutional resources to execute the goal and mostly the experience in the industry is the barrier to executing this activity.

"An approach to closing this gap could be that the two parties work together on the education goals," says [9]. This will increase the capabilities of the universities to meet the industry's expected training goals. Some research studies have designed models to eliminate the barrier of the industry-academia gap. [10] implemented a designed approach to acting as a facilitating strategic tool for better understanding, co-creation facilitated workshops, and prototypes to facilitate the collaboration between industry and academia.

The challenges mentioned above have prompt us to carry out a survey for the NUST, students, and staff, including the industry experts. This was to get a perspective of the students, academic staff and industry experts on their understanding of Data Science in Namibia. The bellow approaches were used for collecting relevant data for the study.

Surveys were used to collect quantitative data about the demographics of the respondents, where they stand with and their whole understanding of data science. Th survey focused on three variables namely age group, gender, current statuses, Programmes of studies and how long after graduation did participants get employed.

The study made use of the thematic data analysis method, including an inductive approach to analyze qualitative data. The International Business Machine Statistical Package for the Social Sciences (IBM SPSS) and inferential analysis was used to analyze quantitative data.

The population of the study was selected based on forty-eight (48) participants. Non-probability-convenience and purposive sampling method was chosen to gather data from the targeted population to get an in-depth understanding of the students' and graduates' knowledge based on the gap and it is less expensive.

III. Survey and Results Discussion

Results from a total number of 44 survey participants were collected and analyzed for both HET and Industry experts – namely Namibia University of Science and Technology and Standard Bank. Looking at 4 different categories to get results and opinions: Studying, Working and studying, Working, HET staff.

A. *Survey Results*

The survey aimed at collecting participants reviews on their understanding of the Data Science skills gap and their understanding of the use of NLP to bridge the skills gap. 65% of the participants have an understanding of both NLP and DS relating to the explanation provided to them in the survey.

TABLE 1. IN YOUR UNDERSTANDING OF DATA SCIENCE – AR THE COURSES OFFERED PRACTICAL OR THEORY?

Category	Respondents (R1-R4) - Studying, Working and Studying			
	R1	*R2*	*R3*	*R4*
Studying	"We have more theoretical lessons than hands-on practical work"	"Academic institutions are sometimes concerned with completing syllabus hence theory rather than practical"	"I will say both. We had statistics and AI in ml we covered topics like machine learning (data problems using) and other those other searching algorithms. I have gone far with it because it's something I am interested in (like I some courses on Udemy and YouTube)"	"Doing MSc by Research. There is no coursework that includes Data Science. However, if I were to pick a research topic on Data Science, I am confident that my supervisor(s) would guide me to readily available data science modules."
Working and Studying	"The industry is more of work or practical experiences. It is vital that during the academic process, students are introduced to what is expected out there"	"It's practically based since most methods to be used will be by extracting data from different systems. Analyze these data and finally present data in a graphical way for better understanding to the business."	"most of these data science are done practically only at universities but not at higher school."	"They entail hands-on analytical expertise and practical decisions."

TABLE 2. THE STATED COURSES OF TE SKILLS GAP

Category	Respondents (R1-R4) - Studying, Working and Studying			
	R1	*R2*	*R3*	*R4*
Working	"Higher institutions should have a memorandum of understanding with working industries to identify skills necessary for improving the existing industries"	"Lack of collaboration between the institutions and industry."	"Lack of organizations refusing students to do internships at their company and the higher education institutions not doing more to help students."	"Lack of practical work given to the students instead of theoretical"
HET staff	"I think lack of strong collaboration between industry and HET"	"Curriculum does not speak to working industries' needs and lack of internship"		

TABLE 3. SHOULD THERE BE A COLABORATION BETWEEN INDUSTRY AND HET?

Category	Respondents (R1-R4) - Studying, Working and Studying			
	R1	*R2*	*R3*	*R4*
Working	"Have a partnership with industries that will have some people from the industry teach in the universities"	"By ensuring that both stakeholders understand the importance of data science"	"By working together and do more of direct skill injection to students."	"Provide internships"
Working and Studying	"Stop using lecturers to teach, rather call-in industry expects to come and show us what they want to be done, and how they want it done. The lecturing approach has become too theoretical, and far removed from the practical. It takes a degree 3 years, by the time you are done, the industry is not using what they taught on campus"	"By making job internship compulsory to all students before completion of any degree (more like a model where lectures need to get direct feedback from the industry)"	"By identifying the skills gap and see how can these be incorporated into the study programs, to enable or prepare graduates for the working industry upon graduating. Skills will base on the skills currently lacking in the Namibian IT market."	"Create a more diverse learning system that teaches learners using different methods tailored to their way of learning as well as embracing unconventional courses."
HET Staff	"It could be done by collaborating with the industry and have their employees take the course or have expertise from the industry to give guest lectures"	""	""	""

Further review was done from a survey for all four participants categories to raise their opinions on what needs to be done to improve this factor of lacking skills of most professions in the industry like designing a framework, a model or any technology that will help bridge the data science skills gap in Namibia. Table 4 illustrates the sample response of four respondents on the impact of the framework to both HET and the Industry.

TABLE 4. HOW NLP CAN BE APPLIED IN RESPONSE TO THE EXISTING SKILLS GAP.

Category	Respondents (R1-R4) - Studying, Working and Studying			
	R1	*R2*	*R3*	*R4*
Studying, HET staff, Working and Studying	"You will need the human factor in this technology"	"None at the moment"	"A model that clearly provides guidance on how academic and real-life working environments are mapped."	"No objections"

B. *Quantitative analysis*

Different instruments were used in the study to collect qualitative data. The data was later analyzed based on five demographics, namely: Age, Gender, Programme of study, how long after graduation to get employed and current status. The five demographics are demonstrated in fig1 below. The analysis was done analysis using descriptive statistics with SPSS statistical software to be able to draw graphs for a clear presentation of the data and the expected results as output. In this case, we performed Normality assessment, Variable ranking, Reliability and Correlation analysis as demonstrated below;

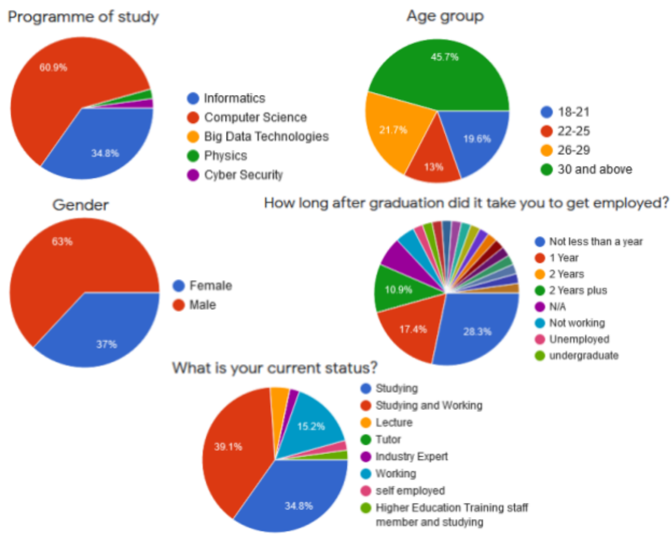

Fig. 1. Demographic Analysis

C. *Normality statistics*

Normality statistics was used to help us identify the level of normality of the data, this is looking at whether the data distribution was normal and had no ambiguity. According to [3] in many cases, statistics inferences require a distribution that is normal or close to normal. When the results are at 0 or close, they state a normal distribution of skewness and access kurtosis. This then determines that your distribution is close to being normal. Based on five variables that help identify the skewness and kurtosis of the variables, Table 6 shows the results of the normality statistics. Based on [11] normal distribution is when the arithmetic skewness is between -2 and 2 and the left and right skew is when the values fall below -2 and above 2. All variables fall withing the normal distribution as displayed in the table. The skewness and kurtosis for Programme of study and Age fall below and above normal distribution, and are insignificant also the presented variable follows a normal distribution. On the other hand, the significance of the variables used in the study is determined by the Mean and the Standard deviation values obtained (Table 5). The variables are considered to be more significant if they have a high mean and law standard deviation, which is the case in the table. Therefore, the variables are significant.

TABLE 5. NORMALITY ASSESSMENT

Variables	N	Mean	Standard Deviation	Skewness	Kurtosis
	Statistics	*Statistics*	*Statistics*	*Statistics*	*Statistics*
Programme	48	1.71	0.683	1.281	3.458
Age	48	2.98	1.176	-0.696	-1.069
Gender	48	0.35	0.483	-0.630	-1.675
Period	48	3.02	1.780	0.015	-1.863
Status	48	2.08	1.200	1.761	3.604
Frame work	48	1.81	0.862	1.107	0.979
Relate	48	1.76	0.790	1.082	1.325
NLP	48	1.95	0.815	0.992	1.203
Skills	48	2.75	0.707	0.404	-0.229
Tools	48	2.16	1.068	0.571	0.778

D. *Reliability analysis*

In this analysis, the study is determining the variable's significance. Data shown in table 6 represents the variable's standard deviation and mean obtained from the data collected. The results tend to be more significant when there is low standard deviation and high mean. In this study, the method used for each variable in the survey were based on how best they disagree or agree or how high or low they rate the design of the framework and the skills gap existence. The results are placed in the analysis as, strongly agree=1; Agree=2; Disagree=3; Strongly Disagree. For high to low: Highest=1; High=2; Low=3; Lowest=4. The table shows a normal distribution which is rated as significant.

E. *Reliability analysis*

The reliability of the study is measured based on the Landis rating scale in table 6 from Cronbach's Alpha approach. Cronbach's Alpha is referred to as a measure to show the internal consistency or the reliability of the data [8]. The reliability of the data is often measured for its collection consistency [11].

TABLE 6. RATING SCALE [11]

Scales	Description
0.81-0.90	Almost Perfect
0.61-0.80	Substantially Reliable
0.41-0.60	Moderate Reliable
0.21-0.40	Fairly Reliable
0.00-0.20	Slightly Reliable

TABLE 7. CRONBACH'S APPROACH FOR MEASURING RELIABILITY ANALYSIS

Cronbach's Alpha	Cronbach's Alpha based on Standardized items	No. of Items
0.910	0.833	49

Table 7 displays the Cronbach's Alpha calculation of 0.910 and Cronbach's Alpha based on standardized items of

0.833 from the Total of 49 items. The difference between the two is Cronbach's alpha uses the covariances of the items while Cronbach's Alpha based on standardized items uses the correlations of the items (IBM, 2020). Therefore, comparing the results to Table 7, the study data and collection methods fall under the range of 0.81-0.90, which is referred to as "*Almost perfect*". This means, the collection methods and data of the study are reliable.

TABLE 8. CORRELATION ANALYSIS

Variables		Programme	Age	Gender	Period	Status	Framework	Relate	NLP	Skills	Tools
Programme	Pearson Correlation	1	-0.326*	-0.003	-0.170	-0.004	-0.065	-0.057	-0.114	-0.477	0.199
	Sig. (2-tailed)		0.024	0.896	0.248	0.977	0.681	0.720	0.483	0.232	0.414
	N	48	48	48	48	48	48	48	48	48	48
Age	Pearson Correlation	-0.326*	1	0.013	-0.203	0.408**	0.342*	0.220	0.213	0.103	-0.361
	Sig. (2-tailed)	0.024		0.929	0.166	0.004	0.026	0.161	0.187	0.808	-0.128
	N	48	48	48	48	48	48	48	48	48	48
Gender	Pearson Correlation	-0.003	-0.013	1	-0.231	0.131	0.040	0.086	0.108	0.0	-0.027
	Sig. (2-tailed)	0.986	0.929		0.114	0.373	0.804	0.587	0.505	1.0	0.913
		48	48	48	48	48	48	48	48	48	48
Period	Pearson Correlation	-0.170	-0.203	-0.231	1	-0.349*	-0.126	0.060	-0.031	0.323	-0.331
	Sig. (2-tailed)	0.248	0.166	0.114		0.015	0.426	0.705	0.851	0.434	0.166
	N	48	48	48	48	48	48	48	48	48	48
Status	Pearson Correlation	-0.004	0.408**	0.131	-0.349*	1	0.070	-0.067	0.060	0.0267	0.0
	Sig. (2-tailed)	0.977	0.004	0.373	0.015		0.660	0.676	0.711	0.522	0.0
	N	48	48	48	48	48	48	48	48	48	48
Framework	Pearson Correlation	-0.065	0.342*	0.040	-0.126	0.070	1	0.648**	0.600**	0.520	-0.195
	Sig. (2-tailed)	0.681	0.026	0.804	0.426	0.660		0.0	0.0	0.231	0.453
	N	48	48	48	48	48	48	48	48	48	48
Relate	Pearson Correlation	-0.057	0.220	0.086	0.060	-0.067	0.648**	1	0.493**	0.496	-0.213
	Sig. (2-tailed)	0.720	0.161	0.587	0.705	0.675	0.0		0.001	0.257	0.411
	N	48	48	48	48	48	48	48	48	48	48
NLP	Pearson Correlation	-0.114	0.213	0.108	-0.031	0.060	0.600**	0.496**	1	0.365	-0.177
	Sig. (2-tailed)	0.438	0.187	0.505	0.851	0.711	0.0	0.001		0.477	0.512
	N	48	48	48	48	48	48	48	48	48	48
Skills	Pearson Correlation	-0.477	0.103	0.0	0.323	-0.267	0.520	0.496	0.365	1	0.0
	Sig. (2-tailed)	0.232	0.808	1.000	0.434	0.522	0.231	0.257	0.477		
	N	48	48	48	48	48	48	48	48	48	48
Tools	Pearson Correlation	0.199	-0.361	-0.027	-0.331	0.0	-0195	-0.213	-0.177	0.0	1
	Sig. (2-tailed)	0.414	0.128	0.913	0.166	0.0	0.453	0.411	0.512		
	N	48	48	48	48	48	48	48	48	48	48
*. Correlation is significant at the 0.05 level (2-tailed).					**. Correlation is significant at the 0.01 level (2-tailed).						

F. Correlation Analysis

Table 8 displays the results from analyzing selected variables for the correlated coefficient. Looking at table 8, it shows that the correlation coefficient between Age and Programme of study is -0.326* which shows a negative correlation – meaning the age does not really restrict one to do a specific programme of in HET. There is also a strong positive correlation between "NLP" and "designing the Framework/Model" feature ("participants with a great understanding of NLP and how it works tend to recommend the design of the framework or model") that is 0.600**. The programme of study from the demographic section displays that Computer science is the Programme with the most participants in the survey questionnaire with 58.3% based on the google form survey data analysis. This correlation is again significant at the 0.05 level (2-tailed).

There is also a strong correlation coefficient of 0.408** between current status and Age. Skills comparison of the framework and the framework Builds a supportive Lecturer and student community feature correlate with a strong positive correlation .648**. Therefore, it can be concluded that Current status and Age and Designing the framework strongly correlate and the age may also defines how long it takes for one to get hired in the DS job. The correlation is significant at 0.01 level (2-tailed).

CONCLUSION AND FUTURE STUDY

Overall, the survey focused on understanding Data Science and NLP, the existing skills gap, the tools needed to qualify for a Data Science profession and innovation in general. Participants of the survey showed a great understanding of data science and how they would dearly want the education system to work in their favor. Most participants ideally agree to the collaboration of HET and the industry to improve the skills offered and prepare students before employment. Most participants do not object to the use of NLP to design a system in favor of bridging the gap. Therefore, some stated that it would be better if this system will be a model that clearly provides guidance on how academic and real-life working environments are mapped and that it should still consider the human factor in the NLP technology. Participants from the industry experts also emphasized that the most cause of the existing skills gap is that there is no collaboration between HET and the industry organizations, HET seem to not often review their curriculums mased on the rapid changes between HET and the industry. The next step in this regard, following recommendations from participants, is to design a framework model that will act as a recommender system to display and compare the most needed skills and tools in the field of science. Academia can use this as a window of measurement when restructuring the curriculum.

REFERENCES

[1] Vahid , G., Görkem , G., Eray , T., Cagatay , C., & Michael , F. (2020). Closing the gap between software engineering education and industrial needs. IEEE software, 68-77.

[2] Kollia, I., & Siolas, G. (2017). Using the IBM Watson cognitive system in educational contexts. 2016 IEEE Symposium Series on Computational Intelligence (SSCI). Athens, Greece: IEEE.

[3] Cárdenas-Navia, s., & Fitzgerald, B. K. (2015). The broad application of data science and analytics: Essential tools for the liberal arts graduate. Change: The magazine of higher learning, 25-32.

[4] Hwang, J. (2020). NLP visualizations for clear, immediate insights into text data and outputs. Retrieved from plotly: https://medium.com/plotly/nlp-visualisations-for-clear-immediate-insights-into-text-data-and-outputs-9ebfab168d5b

[5] Education, I. C. (2020). What is Natural Language Processing? Retrieved from IBM: https://www.ibm.com/cloud/learn/natural-language-processing.

[6] Pulipaka, S. K., Kasaraneni, C. K., Vemulapalli, V. N., & Kosaraju, S. S. (2019). Machine Translation of English Videos to Indian Regional Languages using Open Innovation. 2019 IEEE International Symposium on Technology and Society (ISTAS). Medford, MA, USA: IEEE.

[7] Shaanika, I., & Iyamu, T. (2018). Health informatics curriculum development for teaching and learning. Springer Link, 18.

[8] Valstar, S., Krause-Levy, S., Macedo, A., Griswold, W. G., & Porter, L. (2020). Faculty views on the goals of an undergraduate CS education and the academia-industry gap. SIGCSE '20: Proceedings of the 51st ACM technical symposium on computer science education (pp. 577 - 583). Portland, OR, USA: association for computing machinery, New York, NY, United States.

[9] Beckman, K., Neal, C., Khajenoori, S., & Mead, N. R. (1997). Collaborations: Closing the industry–academia gap. IEEE software, 49.

[10] Wallin, J., Isaksson, O., Larsson, A., & Elfström, B.-O. (2014). Bridging the gap between university and industry: Three mechanisms for innovation efficiency. International journal of innovation and technology management, 1-15.

[11] Nakale, B., & Chikohora, E. (2020). Towards A Data-Driven Innovation Business Model in Digital Government Services in Namibia: A Survey. International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (p. 4). IEEE.

[12] Brown, S. (2011). Measures of Shape: Skewness and., (p. 16).

[13] Education, N. C. (2018). 2018 Namibia Higher Education Statistical Yearbook (Nhesy). Windhoek: National Council for Higher Education.

[14] Demchenko, Y., Belloum, A., Los, W., Wiktorski, T., Manieri, A., Brocks, H., . . . Becker, J. (2016). Edison data science framework: A foundation for building data science profession for research and industry. 2016 IEEE international conference on cloud computing technology and science (cloudcom). Luxembourg City, Luxembourg : IEEE.

[15] Gasper, R. E., & Baker, C. M. (2020). Data Science in 2020: Computing, Curricula, and Challenges for the Next 10 Years. Journal of Statistics and Data Science Education.

[16] Goforth, C. (2015). Using and Interpreting Cronbach's Alpha. Retrieved from Research and data servces + Sciences: https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/#:~:text=Cronbach's%20alpha%20is%20a%20measure,of%20scale%20or%20test%20items.&text=Cronbach's%20alpha%20is%20thus%20a,variance%20of%20the%20total%20score.

[17] IBM. (2020). Difference between alpha and alpha for standardized items in Reliability output. Retrieved from IBM Support: https://www.ibm.com/support/pages/difference-between-alpha-and-alpha-standardized-items-reliability-output#:~:text=The%20first%20Cronbach's%20alpha%20employs,is%20often%20false%20in%20practice.

[18] pwc. (2017). Investing in America's Data Science and Analytics Talent. Retrieved from The Business-Higher Education Forum: https://www.bhef.com/publications/investing-americas-data-science-and-analytics-talent

[19] Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). A Python Natural Language Processing Toolkit. Cornell University. Stanford: Cornell University.

[20] Valstar, S. (2019). Closing the academia-industry gap in undergraduate CS. ICER '19: Proceedings of the 2019 ACM conference on international computing education research (pp. 357-358). Toronto, ON, Canada: Association for computing machinery, New York, NY, United States.