Available online at www.alphanumericjournal.com

alphanumeric journal
The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 10, Issue 2, 2022

Received: June 07, 2022
Accepted: July 01, 2022
Published Online: December 31, 2022

AJ ID: 2022.10.01.OR.02
DOI: 10.17093/alphanumeric.1127620
Research Article

Application of a Combined Approach of Text Mining and QFD Methodology Based on Single Valued Neutrosophic Numbers for Efficient Curriculum Design

Sevgi Abdalla, Ph.D. *

Assist. Prof., Faculty of Science and Letters, Department of Statistics, Eskişehir Osmangazi University, Eskişehir, Turkiye, sayhan@ogu.edu.tr

* ESOGÜ Fen Fakültesi, Meşelik Kampüsü Büyükdere Mah. Prof. Dr. Nabi AVCI Bulvarı No:4, 26040 Odunpazarı, Eskişehir, Türkiye

ABSTRACT

In this study, an alternate curriculum design for an undergraduate program of Statistics is suggested carrying out a combined approach of the QFD methodology, text mining techniques under single valued neutrosophic set environment. To capture the employers' expectations from their potential employees, 640 job advertisements, obtained from two of the most important career and job posting sites in Turkey, were analyzed using TF-IDF technique, which is one of the text mining methods. By using single-valued neutrophic set (SVNS) theory in QFD, the technical requirements representing the courses included in the curriculum were found their priorities. Hence, the technical characteristics that play a critical role in evaluating the curriculum quality of the undergraduate program were revealed. In addition, single valued neutrosophic sets have provided a flexible decision-making procedure to improve the quality of individuals' subjective assessments. Consequently, this is expected to be a good reference for researchers working on these issues, both in terms of the proposed approach and the problem addressed.

Keywords: Single Valued Neutrosophic Sets, Quality Function Deployment, TF-IDF Measure, Curriculum Design

1. Introduction

Quality has become one of the critical concepts for companies and individuals in terms of being able to compete and maintain their existence in today's global environment. With the incredible progress in technology, quality improvement efforts in all areas have also gained momentum. As a result of this process, quality management has taken its place as one of the indispensable requirements of sustainable management approach. Improving and improving the quality of education and training services in all areas where the human factor is effective is directly related to social development.

In this context, education is not only a social interest but also a fundamental economic issue (Kelesbayev, 2014). Hanusek et al. (2015) stated that one of the key elements of economic development policies is the improvement of human capital. In particular, the fact that any curriculum educates individuals who are adequately equipped to meet employer expectations is one of the most important factors in improving human capital. For this purpose, designing curriculum of an education program that will improve the technical and social skills of students in a way that meets employers' expectations.

In the literature, there have been some important studies were proposed to handle the curriculum design problem for different education programs by different perspectives. For instance, Köksal and Eğman (1998) and Jnanesh and Hebbar (2008), proposed two different studies to improve the quality of the industrial and mechanical engineering programs, respectively, have prioritized the expectations of students, employees and instructors by weighing and finally, common expectations as customer requirements have found. However, the researchers have not suggested curriculum. Abuzid (2017), in their study, they have worked on determining the right teaching strategies to meet the students' needs studying in the business administration. Gupta et al. (2012) have conducted a study on the determination of effective courses that provide information about quality management to be included in the curriculum of an industrial engineering department. In addition, to these studies, Boonyanuwat et al (2008) and Ünal and Uysal (2014) evaluated courses for industrial engineering education programs have been carried out in their own countries in terms of student expectations. In all these studies mentioned in the literature, quality function deployment methodology has been carried out for quality evaluation process and improvement of a program curriculum. Ünal and Uysal (2014) solved the problem using mixed integer programming approach. Kamvysi et al (2014) have determined student expectations which provide better knowledge and skills they have titled in five different categories and prioritized these categories by combining the Fuzzy-AHP method linear programming approaches in QFD. As a result, it is seen that QFD methodology is preferred as the most effective approach for the curriculum design of any education program.

QFD is a quality improvement tool that converts customer needs and expectations into the technical specifications of a product or service (Erdil and Arani, 2018). The most important advantage of QFD is that it is easily integrated with statistical, mathematical, and artificial intelligence-based algorithms to draw conclusions. Thus, the method still offers a powerful quality assessment platform today without losing

its relevance. With all these advantages, the QFD method requires a decision-making process because of its problem-solving structure. In other words, the method produces results based on the subjective evaluations or preferences of the decision-makers. These outputs may contain uncertainty, contradiction, or bias information (Smithson, 2015).

Recently, various mathematical approaches have been suggested to improve the quality of a decision based on such subjective judgements. Fuzzy set theory, which is one of the most efficient approaches, introduced by Zadeh (1965) (Zadeh, 1965). The approach produces a membership value for an element which belongs to the relevant set using membership function. However, fuzzy set theory is insufficient to provide information that an element does not belong to a set. (Van et al., 2018). For that purpose, neutrophic set theory (NST), which is also derived from fuzzy set theory and includes the uncertainty and falsity information contained in a decision, has been introduced by Smarandache in 1999 (Sodenkamp et al., 2018). Due to the difficulties in applying to real-life problems, Wang et al. (2010) improved the theory and presented to the literature a single valued neutrophic set (SVNS) which provides additional information as possibility that represents uncertainty, imprecise, incomplete, and inconsistent information (Ye, 2014; Pramanik et al., 2018). A neutrosophic set consists of the three membership functions that represent truth, indeterminancy, and falsity situations, respectively. Hence, the approach is one of the hot topics that has been using in decision-making as an alternative approach to fuzzy set theory over the last 5 years.

In this study, an effective and innovative approach to evaluate and to suggest an improved curriculum for an undergraduate program of Statistics in which the text mining methods TF-IDF and single-valued neutrophic set (SVNS) theory are integrated into the QFD method is presented. Thus, the technical characteristics that play a critical role in the evaluation of the curriculum quality of the undergraduate program are prioritized.

2. Material and Method

In this part of the study, the mathematical background of and the techniques and methods that will form the basis of the proposed approach are explained.

2.1. Single Valued Neutrosophic Sets

Some definitions of single valued neutrosophic sets defined by Smarandache (1999) is given in following (Wang et al., 2010; Ye, 2014).

Definition 1 (A *Neutrosophic set):* Let X be a space of objects and a neutrosophic set *A* is described by the three membership functions are called truth ($T_{A(X)}$), indeterminancy ($I_{A(X)}$), and falsity ($F_{A(X)}$), respectively. The form of a SVNS is shown in Eq 1.

$$\boldsymbol{A} = \{\langle \boldsymbol{x}, \boldsymbol{T_A}(\boldsymbol{x}), \boldsymbol{I_A}(\boldsymbol{x}), \boldsymbol{F_A}(\boldsymbol{x})\rangle : \boldsymbol{x} \in \boldsymbol{X}\}. \quad (1)$$

These functions are denoted as $\boldsymbol{T_A}(\boldsymbol{x}): \boldsymbol{X} \rightarrow]^{-}\boldsymbol{0}.\boldsymbol{1}^{+}[$; $\boldsymbol{I_A}(\boldsymbol{x}): \boldsymbol{X} \rightarrow]^{-}\boldsymbol{0}.\boldsymbol{1}^{+}[$; $\boldsymbol{F_A}(\boldsymbol{x}): \boldsymbol{X} \rightarrow]^{-}\boldsymbol{0}.\boldsymbol{1}^{+}[$, **respectively**, and satisfying the formula given in Eq. 2.

$$\boldsymbol{0}^{-} \leq \boldsymbol{sup T_A}(\boldsymbol{x}) + \boldsymbol{sup I_A}(\boldsymbol{x}) + \boldsymbol{sup F_A}(\boldsymbol{x}) \leq \boldsymbol{3}^{+} \quad (2)$$

Definition 2 *(a single valued neutrosophic set):* a SVNS (*A*) on *X* is described by the three membership functions such as truth-membership function $\boldsymbol{T_A}$, indeterminacy-membership function $\boldsymbol{I_A}$ and falsity-membership function $\boldsymbol{F_A}$, if *x* represents a single valued independence variable. Each object *x* in *X* belongs to subset of real numbers [0, 1]. SVNS (*A*) can be written for continuous cases as in Eq. 3 and for discrete cases as in Eq. 4.

$$\boldsymbol{A = \int T(x), I(x), F(x)/x\ ,\ \ x \in X} \tag{3}$$

$$\boldsymbol{A = \sum_{i=1}^{n} \langle T(x_i), I(x_i), F(x_i)\rangle / x_i\ ,\ \ x_i \in X} \tag{4}$$

Definition 3 *(Basic operators of SVNS*): All classical set operators (i.e., complement, intersection, union, containment, difference, cartesian product) are also applied in SVNS theory.

Let *A* and *B* be two single-valued neutrosophic numbers which are denoted by $\boldsymbol{O = \langle T_A, I_A, F_A\rangle}$ and $\boldsymbol{O_B = \langle T_B, I_B, F_B\rangle}$, respectively. The following arithmetic relations are given in the following forms.

$$\boldsymbol{O_A \oplus O_B = \langle T_A + T_B - T_A T_B,\ I_A I_B,\ F_A F_B\ \rangle} \tag{5}$$

$$\boldsymbol{O_A \otimes O_B = \langle T_A T_B,\ I_A + I_B - I_A I_B,\ F_A + F_B - F_A F_B\ \rangle} \tag{6}$$

$$\boldsymbol{\alpha O_A = \langle 1 - (1 - T_A)^{\alpha}, I_A^{\alpha}, F_A^{\alpha}\rangle\ \ for\ \alpha > 0} \tag{7}$$

$$\boldsymbol{O_A^{\alpha} = \langle O_A^{\alpha}, 1 - (1 - I_A)^{\alpha}, 1 - (1 - F_A)^{\alpha}\rangle\ \ for\ \alpha > 0} \tag{8}$$

Definition 4. The score value $\boldsymbol{S(O_A)}$ of a SVNN $\boldsymbol{O_A = \langle T_A, I_A, F_A\rangle}$ are calculated through the following formula.

$$\boldsymbol{S(O_A) = (2 + T_A - I_A - F_A)/3} \tag{9}$$

2.2. Text Mining Methods and Programming Language

In this part of the study, TF-IDF technique used in the text analysis are explained and a brief information about the Python software is given.

2.2.1. TF-IDF (Term Frequency- Inverse Document Frequency) Technique

TF-IDF is a statistical -based machine learning and information retrieval method used to extract confidential information contained in text documents (Bafna et al., 2016). TF-IDF was developed from IDF, which was proposed by Sparck Jones as a query term with heuristic intuition (Zhang et al., 2011).

The frequency (TF) of a term represents a measure of how important a word in the text can be. Therefore, the importance of the word in question is based on determining how often this word occurs in a document. However, a document also contains words that are repeated many times but may not be important. For example, these words included in Turkish texts are probably "and", "or" "that" etc. such as words. In the analysis of text mining, the approach of removing frequently used words such as conjunctions and adjectives from the text is adopted.

The term frequency ($\boldsymbol{tf(t, d)}$) of a word is calculated by the formula given in Equation 1 below. Here, the relevant word is represented by *t* and document by *d.* It is obtained

by the ratio of the number of words in the text to the number of all words in the text.$\boldsymbol{f_{t,d}}$

$$\boldsymbol{tf(t,d)} = \frac{\boldsymbol{f_{t,d}}}{\boldsymbol{total\ term\ number}} \quad (10)$$

Sometimes the logarithmic scaled frequency ($\mathbf{log(1 + tf(t,d))}$ is also used in studies (Kınık and Güran, 2021). Another approach is the reverse document frequency (IDF) method, which reduces the weight of commonly used words in a document and increases the weight of words that are not used much (Ramos, 2003). This method is used with term frequency to calculate the term's frequency (TF-IDF), adjusted for how often a term is used. The reverse document frequency for any term is defined as:

$$\boldsymbol{idf(t,d)} = \mathbf{log}(\frac{\boldsymbol{N}}{|\{\boldsymbol{d \in D : t \in d}\}|}) \quad (11)$$

Here, the IDF value is obtained by dividing the total number of documents N by the number of documents containing the t-word. The $\{\boldsymbol{d \in D : t \in d}\}$ IDF value provides information about the relative rarity of a word (Zhou, 2022). In other words, if the IDF value approaches 0, that word indicates a common or ordinary word in all documents, if its value approaches 1, that word is an important word that is not included in every document.

In addition, by multiplying both values ($\boldsymbol{tf(t,d) * idf(t,d)}$), the TF-IDF value is obtained, which represents the relative importance of a word and points to critical attention-grabbing words in keyword extraction. In addition, a general text summary is obtained by determining the number of sentences by the researcher/programmer by looking at the frequency of the sentences in which the keywords are mentioned.

2.2.2. Python Programming Language

Python programming language is a high-level, dynamically typed language that is among the most popular general-purpose programming languages. It is among the fastest growing programming languages in the world and is used by software engineers, mathematicians, data analysts, scientists, network engineers, students, and accountants. The biggest advantage of the Python language is that it offers a huge collection of standard libraries for big data projects such as machine learning, image processing, scientific computing.

Considering all these advantages, in this study, all text mining analyzes were performed using the NLTK (Natural Language ToolKit) library in Python 3 software in all text mining analyzes. To avoid any loss of information due to the Turkish grammar deficiencies of the relevant libraries of the programming language, the data processing and analysis process was carried out over the English text.

2.3. Quality Function Deployment

Quality Function Deployment (QFD) was introduced by Yoji Akao in Japan in 1966 which was initially theoretical work, then the methodology was put into practice in 1972 at the Kobe shipyard of Mitsubishi Heavy Industries Limited with the participation of Dr. Mizuno and Dr. Furukawa (Dean, 1995). After 1978, Toyota reduced product development costs by 61%, shortened the product development process and eliminated rust problems (Abdul-Rahman et al., 1999). Recently, QFD has provided a methodology which aims to improving design quality to meet customer requirements and transforms those requirements into design objectives

and major quality assurance points to be used during production (Erdil and Arani, 2018). Consequencly, QFD is a convenient method that acts as a bridge in establishing a good communication between customers and manufacturer by accomplishing this transformation. Reflecting the requirements of the customers correctly to the product means ensuring customer satisfaction, which is one of the objectives of the companies.

QFD methodology consists of six main phases which are listed as follows (Ünal and Uysal, 2014).

1. To catch Voice of Customer (WHATs- What expectations of customers)
2. Find out relative importance of customer needs (How important these whats)
3. Determine technical characteristics of a product or a service (HOWs- How these technical requirements meet customers need)
4. Relationships matrix (Connections between customer requirements and technical characteristics of a product)
5. Prioritize the technical characteristics of a product
6. Benchmarking (Compare a company with its competitors in terms of their performances)

3. Proposed Approach based on QFD for Curriculum Design

3.1. Data and Preprocessing

In this study, an alternative curriculum design for undergraduate program of Statistics is suggested by carrying out a combined approach of the QFD methodology by revealing the employers' expectations from their potential employees. For this purpose, 640 job ads obtained from two of the most important career and job posting sites in Turkey. Therefore, the raw data was collected to catch the voice of employers. Then, data preprocessing procedure has implemented using Python programming language.

Considering the inadequacies of the Python programming language libraries in terms of the Turkish language, the ads used in the study were translated into English. Thus, it was ensured that the quality outputs were produced with minimal loss of information. Then, the ads in English language were arranged as a single text and treated as a data set of the study.

In the next step, the English text was purged of punctuation, numbers, digits, and symbols. Thus, the words and concepts in the text have been made simpler. The words expressed as main text data by removing the ineffective words (stopwords) that mentioned above from the text (Patil and Atique, 2013). These ineffective words are often used in a language such as conjunctions, adverbs, auxiliary verbs, and pronouns. Thus, it becomes easier to reveal the words related to the subject to be emphasized in the text. The words that have different suffixes but have the same root are prevented from being perceived as different words by the program.

3.2. Proposed Approach for Curriculum Design of a Statistics Undergraduate Program

In this part of the study, the implementation steps of the proposed combined approach of QFD for curriculum design are explained in detail and the results obtained are discussed.

3.2.1. Determining the Employers' Expectations (Whats) Using Text Mining Methods

The first step in QFD is to identify who the customers are as the target group. This step is the most critical part of the methodology to producing the right product that meets customer needs. Customer needs can be determined in different ways such as conducting surveys, considering requests, complaints, and suggestions from customers by phone, fax, e-mail, social media platforms or e-commerce web sites are the most frequently applied methods.

In the study, companies which have job ads on career sites were identified as customers and the expectations of employer companies from Statistics graduates were determined by analyzing the data set using text mining methods as TF-IDF technique.

Initially, significant words in text representing employer expectations were revealed from job ads. The process of obtaining these words is expressed as keyword extraction in the text mining literature.

As a result, employer expectations obtained for this study, expressed as customer needs, are given in Table 1. The TF-IDF method ranks the words according to the frequency of their inclusion in the text. This measurement provides the importance values of the words, which were assigned as the relative weights of the employers' expectations in the quality house and are demonstrated in Table 1. In addition, the relative weights of each employer requirement were calculated. Thus, the actual contributions of the requirements are determined by their degree of importance.

Groups of the Skills	Codes	Employers' expectations	Absolute Weights of expectations	Relative weights of expectations
Social Skills	CR1	Teamwork	11,29	0,118
	CR2	Communication skills	13,42	0,140
	CR3	Analytical thinking	4,18	0,044
	CR4	Foreign Language	4,91	0,051
Software abilities	CR5	Database management systems (SQL)	14,95	0,156
	CR6	MS Office (Excel)	8,10	0,085
	CR7	Programming in R-Phython-SAS	7,92	0,083
	CR8	Statistical Packages	12,37	0,129
Professional/Job Skills	CR9	Optimization	5,95	0,062
	CR10	Quality Management	5,44	0,057
	CR11	Reporting	6,45	0,067
	C12	Project management	12,06	0,126

Table 1. Employers' expectations as customer needs and their importance weights

When the results in Table 1 are examined, companies in Turkey Statistics graduates to have gained skills in Database Management Systems with a significance level of 15.6 %. The development of communication skills of graduates plays a critical role with score value 14 % in meeting the expectations of companies. These characteristics are followed by the ability to use statistical package programs (12.9%), project management (12.6%) and teamwork adaptation skills (11.8%),

respectively. In terms of employer company expectations, analytical thinking ability (4.4%) was determined as the least important criteria.

3.2.2. Determining the Technical Characteristics (Engineering Characteristics)

Technical requirements are called engineering characteristics, which the customers cannot determine or aware that must be considered at the beginning of production process. These requirements could be technical or legal properties of a product or a service. Legal requirements are mostly determined as safety requirement, legal regulations, quality levels, product standards.

In this study, the compulsory and elective courses that will support the development of the basic skills and technical equipment of the graduates of the undergraduate program that will meet the employer expectations at the best level are determined as technical requirements. Totally 12 technical technical characteristics were determined by brainstorming method with a group of heads of departments related to statistical science. In the determination of technical requirements, it is aimed to create the ideal course curriculum that will meet the expectations. These technical requirements represent the critical courses and trainings that should be included in the content of a curriculum. In Table 2, the requirements to be located on the roof of the House of Quality are summarized.

T1- Applications of Statistical Techniques	T7- Business
T2- Operational Research and Decision Making	T8- Quality management
T3-Data Mining	T9- Oral presentation skills
T4-Advanced Statistical Computing (SAS/R/Phyton)	T10-Undergraduate research
T5- Business analytics with Excel	T11-Business Communication and Leadership
T6- Finance and Economics	T12-Business English

Table 2. Technical requirements of the curriculum design for Statistics program.

3.2.3. Neutrosophic Relationship Matrix

The relationship matrix contains the correlations between customer needs and the technical characteristics of a product or service. To determine these relationships, experts, or decision maker -in this study program designers- questioned to assign the degree of the relationships between pairs of requirements. In the study, four levels Likert scale is suggested evaluation the relationships between both requirement groups. The degrees of relationships are represented with linguistic terms as weak (1), fair (3), good (5) and excellent (9). Then, these linguistic terms are converted to a single valued neutrosophic numbers using the scale which is shown in Table 3.

Linguistic terms	**Symbols**	**Crisp numbers**	**A single valued neutrosophic numbers**
weak	Δ	1	<0.1;0.8;0.9>
Fair	○	3	<0.3;0.75;0.7>
Good	◎	5	<0.70;0.15;0.2>
Excellent	●	9	<0.9;0.1;0.1>
Neutral	-	0	<0,5;0,5;0,5>

Table 3. Linguistic terms and the single valued neutrosophic numbers of the relationships

The relationship matrices which include two decision makers' evaluations that are representing the correlations between employers' requirement s and technical characteristics of suggested undergraduate curriculum separately are given in Table 4 (a) and Table 4 (b).

(a) The Relationship Matrix by Decision Maker 1													
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	
C1									❍	-	◉	◎	
C2									◉	❍	◉	◉	
C3	◉	◉	◉	◉	◉	❍			❍	◉	◎		
C4												◉	
C5			◉	◉									
C6	❍				◉	◎			◉	◎			
C7	◉		◉	◉	◎					◎			
C8	◉	◎	-	◎				❍		◉			
C9		◉			◎		❍						
C10					◎		❍						
C11	△	△	△	△	◉			◉		◉			
C12	◎	◉	◎	❍			❍	◉		◉			
(b) Relationship matrix by Decision Maker 2													
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	
C1									◎	◎	◉	◉	
C2									◉	△	◉	◉	
C3	❍	◉	◎	❍	❍	❍			❍	◉	◉		
C4												◎	
C5			◉	◎									
C6	△				◉	◎			◉	◎			
C7	❍		◉	◉	◉					◎			
C8	◉	◎	❍	-				❍		◉			
C9		◉			◎		❍						
C10					❍		❍						
C11	◉	-	◉	△	◉			◎		◉			
C12	❍	◎	△	-			◎	◎		◎			

Table 4. The relationship matrices by Decision Makers

The correlations that score values between the two requirements is calculated in following. Initially, an aggregated single valued neutrosophic number is found using any centered tendency measure as arithmetic, geometric or any average measure. In the study, geometric average operator is suggested for calculating the aggregated single valued neutrosophic number. For example, let decision makers' evaluations for analytical thinking and applications of statistical methods be *fair* and *excellent*, respectively. The corresponded single valued neutrosophic numbers are <0.3;0.75;0.7> and <0.9;0.1;0.1>. For each membership function, the aggregation value is computed as follows.

$T_{CR3-TR1}^{Agg.} = [(0{,}3) * (0{,}9)]^{1/2} = 0{,}5196$ for the truth function,

$I_{CR3*TR1}^{Agg.} = 1 - ((1 - 0{,}75)^{\frac{1}{2}} * (1 - 0.1)^{\frac{1}{2}}) = 0.5256$ for the indeterminacy function, and
$F_{CR3*TR1}^{Agg.} = 1 - ((1 - 0.7)^{\frac{1}{2}} * (1 - 0.1)^{\frac{1}{2}} = 0{,}4804$, for the falsity function.

Next, the score value (S_{corr}) represents the correlation value between analytical thinking and applications of statistical techniques is determined as 0, 5045 using the formula is given in Eq. 23. All results obtained for pairs of requirements are presented in Table 5.

3.2.3. Determining the Relative Weights of the Technical Requirements

In this part of the study, the relative weights of the courses and trainings that ought to be included in the curriculum to educate graduates who are equipped to meet the expectations of the employer have been determined. For this purpose, the relative weights of the customer's requirement representing each employer's expectation

and the correlation values with the technical requirement in return were multiplied and the sum was obtained. Thus, their absolute weight is determined for each of the technical characteristics. After that, the general relative weight of each technical requirement was calculated and all QFD results were presented in the quality house given in Table 6.

4. Results

According to the results of the quality assessment given in Table 5, it is seen that the most important professional course that should be included in the curricula of the Department of Statistics on the training of graduates who will meet the expectations of the employer is the course of graduate research (15.7%). This course, which is related to the solution of a real-life problem from a statistical point of view before the students graduate in their senior year, plays a critical role in terms of experiencing real life. In addition, Data mining and Advanced Statistical Computing courses have the second priority rating with 10.4%. These courses are followed by Business analytics with excel (0.92) and Operations Research- Decision making (9.2%) courses, respectively. it has been proven that these courses, which are determined to be critically important, are the ones that should be given priority in the curriculum.

CRs	TR1	TR2	TR3	TR4	TR5	TR6	TR7	TR8	TR9	TR10	TR11	TR12
CR1									0,471	0,626	0,900	0,840
CR2									0,900	0,194	0,900	0,900
CR3	0,505	0,900	0,840	0,505	0,783	0,283			0,283	0,900	0,840	
CR4												0,840
CR5			0,900	0,840								
CR6	0,194				0,900	0,783			0,900	0,783		
CR7	0,505		0,900	0,900	0,839					0,783		
CR8	0,900	0,783	0,376	0,626				0,283		0,900		
CR9		0,900			0,783		0,018					
CR10					0,571		0,016					
CR11	0,346	0,258	0,346	0,133	0,900			0,840		0,900		
C12	0,471	0,840	0,323	0,376			0,059	0,840		0,840		
Abs. Weigh	0,279	0,320	0,365	0,365	0,322	0,079	0,228	0,199	0,271	0,554	0,269	0,268
Relative Weights	0,079	0,091	0,104	0,104	0,092	0,022	0,065	0,057	0,077	0,157	0,076	0,076

Table 5. Results for Curriculum design in the House of Quality

According to the table, employers have minimum expectations from the graduates of Statistics in terms of finance and economics knowledge. Therefore, it seems that it is sufficient to support candidates of employees who have graduated from the relevant department with supplementary courses such as Applications of Statistical Techniques (7.9%), Oral presentation skills (7.7%), Business English (7.6%). However, the fact that candidates have the ability to use specific methods and computer software of artificial intelligence and statistical science has proven to be the most distinctive factors in meeting the expectations of the employer. In addition, graduates who are gained business (6.5%) and quality management (5.6%) knowledge are also more advantageous in terms of admission to work.

5. Discussion and Conclusions

In today's competitive environment, the improved social skills of employees in terms of employers have been proven to be very effective in hiring. Moreover, it is not even a job for graduates with the ability to use database management systems and project management to stay one step ahead of their competitors. In today's global world, artificial intelligence science dominates all areas without question, making employers more willing to employ graduates who closely follow technology to increase their competitiveness. Of course, the ability of all these social and professional skills as well as the methods and techniques required by statistics science to easily apply in all areas offers highly flexible business opportunities to graduates.

According to the results, graduates with the equipment to meet their employer's expectations are expected to closely follow the developments in the world, especially when reviewing the course content, to improve themselves in machine learning – artificial intelligence. Consequently, these courses, are represented by the technical requirements with higher relative importance, will provide advantages to the students who graduate from the Statistics undergraduate program by developing their skills in social communication and current artificial intelligence technologies.

Besides, an efficient hybrid approach has been proposed by combining text mining techniques and QFD under SVNS environment to enrich the curriculum of Statistics undergraduate program in this study. The approach has produced valuable outputs for improving the quality of undergraduate program. In addition, when the literature was examined, it was observed that there was a limited number of studies that suggested effective methods for a new curriculum design. All methods used in this study are integrated with single valued neutrosophic set theory, providing an innovative decision-making approach. And there was no study in the world literature that looked at curriculum design from that point of view. This study is expected to shed light on researchers working on these issues, both in terms of the proposed approach and the problem addressed.

References

Abdul-Rahman, H., Kwan, C. L., & Woods, P. C. (1999). Quality function deployment in construction design: application in low-cost housing design. International Journal of Quality & Reliability Management.

Abuzid, H. F. T. (2017). Applying QFD tools for quality improvements in curriculum design and teaching strategies to meet with the customer (learner) needs. Journal of Engineering and Applied Sciences, 12 (3), 684-690.

Bafna, P., Pramod, D., & Vaidya, A. (2016, March). Document clustering: TF-IDF approach. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 61-66). IEEE.

Boonyanuwat, N., Suthummanon, S., Memongkol, N., & Chaiprapat, S. (2008). Application of quality function deployment for designing and developing a curriculum for Industrial Engineering at Prince of Songkla University. Songklanakarin Journal of Science & Technology, 30(3).

Dean, E. B. (1995). Quality Function Deployment. Design for Competitive Advantage, World.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. European Economic Review, 73, 103-130.

Erdil, N. O., & Arani, O. M. (2018). Quality function deployment: more than a design tool. International Journal of Quality and Service Sciences.

Gupta, R., Gupta, S., & Nagi, K. (2012). Analysis & designing an engineering course using QFD. International Journal of Modern Engineering Research, 2(3), 896-901.

Güran, A., & Kınık, D. (2021). TF-IDF ve Doc2Vec Tabanlı Türkçe Metin Sınıflandırma Sisteminin Başarım Değerinin Ardışık Kelime Grubu Tespiti ile Arttırılması. Avrupa Bilim ve Teknoloji Dergisi, (21), 323-332.

Jnanesh, N. A., & Hebbar, C. K. (2008, October). Use of quality function deployment analysis in curriculum development of engineering education and models for curriculum design and delivery. In Proceedings of the world congress on engineering and computer science (pp. 22-24).

Kamvysi, K., Gotzamani, K., Andronikidis, A., & Georgiou, A. C. (2014). Capturing and prioritizing students' requirements for course design by embedding Fuzzy-AHP and linear programming in QFD. European Journal of Operational Research, 237(3), 1083-1094.

Kelesbayev, D. (2014). Türk Dünyasinin Eğitim Sistemindeki Ortak Mesele: Kalite. Uluslararası Türkçe Edebiyat Kültür Eğitim (TEKE) Dergisi, 3 (2), https://dergipark.org.tr/tr/pub/teke/issue/12850/155765

Köksal, G., & Eğitman, A. (1998). Planning and design of industrial engineering education quality. Computers & industrial engineering, 35(3-4), 639-642.

Patil, L. H., & Atique, M. (2013, February). A novel approach for feature selection method TF-IDF in document clustering. In 2013 3rd IEEE international advance computing conference (IACC)(pp. 858-862). IEEE.

Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning (Vol. 242, No. 1, pp. 29-48).

Pramanik, S., Dalapati, S., Alam, S., Smarandache, F., & Roy, T. K. (2018). NS-cross entropy-based MAGDM under single-valued neutrosophic set environment. Information, 9(2), 37.

Smarandache F. A Unifying Field in Logics. Neutrosophy: Neutrosophic Probability, Set and Logic. Rehoboth: American Research Press; 1999.

Sodenkamp, M. A., Tavana, M., & Di Caprio, D. (2018). An aggregation method for solving group multi-criteria decision-making problems with single-valued neutrosophic sets. Applied Soft Computing, 71, 715-727.

Smithson, M (2015) Probability judgments under ambiguity and conflict. Frontiers in Psychology. 6 674. doi: 103389/fpsyg201500674

Ünal, Y. Z., & Uysal, Ö. (2014). A new mixed integer programming model for curriculum balancing: Application to a Turkish university. European Journal of Operational Research, 238(1), 339-347.

Van LH, Yu VF, Dat LQ, Dung CC, Chou S-Y, Loc NV (2018) New integrated quality function deployment approach based on interval neutrosophic set for green supplier evaluation and selection Sustainability. 10 (3):838. https://doi.org/10.3390/su10030838

Wang, H., Smarandache, F., Zhang Y. and Sunderraman, R., (2010). Single valued Neutrosophic Sets, Multi-space and multi-structure, 4, 410-413.

Ye, J. (2014). Single valued neutrosophic cross-entropy for multicriteria decision making problems. Applied Mathematical Modelling, 38(3), 1170-1175.

Zadeh, LA. Fuzzy sets. Inf Control. 1965; 8: 338- 353.

Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF* IDF, LSI and multi-words for text classification. Expert Systems with Applications, 38(3), 2758-2765.

Zhou, H. (2022). Research of Text Classification Based on TF-IDF and CNN-LSTM. In Journal of Physics: Conference Series(Vol. 2171, No. 1, p. 012021). IOP Publishing.