THE TEACHING OF MATHEMATICS
2020, Vol. XXIII, 2, pp. 90–101

DATA SCIENCE FOR NOVICE STUDENTS: A DIDACTIC APPROACH TO DATA MINING USING NEURAL NETWORKS

Djordje M. Kadijevich

Abstract. This paper presents a way to introduce data science to college (business) students. To this end, data mining using neural networks may be practiced. After briefly clarifying the main differences between data science and data mining, the work with neural networks is examined in detail. The examination deals with the features and affordances of this work, as well as its expected challenges with possible reasons. The paper ends with a number of implications for practice, teacher education, and research.

MathEduc Subject Classification: K85, P85

MSC Subject Classification: 97K80, 97P80

Key words and phrases: Neural networks; data mining; data science; college education.

1. Introduction

Various initiatives worldwide call for equipping student with data analytics and data science skills to have them better prepared for the jobs of the future. For example, Asia Pacific Economic Cooperation Agency developed a list of Data Science and Analytics (DSA) Competencies, including: enhanced skills in data presentation and visualization; versatile applications of data analytics methods; computational thinking and use of algorithms [1]. It is predicted that job candidates with DSA competencies will soon be preferred by most employers in developed economies (for this preference, see `http://www.bhef.com/sites/default/files/bhef_2017_investing_in_dsa.pdf`).

Data science is, in brief, concerned with preparing data and finding useful information in data by using exploratory data analysis or other (advanced) mathematical/statistical methods. To extract such information, several fields are often combined, including database processing, programming, statistics, optimization, data visualization, pattern recognition, data mining, and machine learning. Data scientists are thus concerned with creating various data products, such as valuable datasets and novel knowledge discovery applications that outperform conventional data processing software (e.g., [3]).

A central field of data science is data mining. This field basically deals with uncovering relationships and patterns buried in large datasets, and such data insights are usually used to make informed decisions in order to maintain the enterprise in question [26]. To describe data mining, the so-called SEMMA cycle may be used. Its key stages are: *Sample*–sampling data and creating partitioned datasets for fitting, assessing and generalizing from model; *Explore*–applying exploratory statistical and visualization techniques to search for trends, anomalies, and clusters in the data set; *Modify*–selecting, creating and transforming the variables; *Model*–searching for variables that predict a desired outcomes reliably; *Assess*–evaluating the model to find out how well it performs [19].

As workplace contexts change, it is reasonable to expect that students will increasingly encounter key elements of data science, at least at an introductory level (e.g., [12]). To practice data science at this level, data visualization using interactive charts may be applied [10]. Regarding data mining, work with neural networks (NNs) may be suitable.

NNs represent one of a number of data mining approaches, including exploratory data analysis, liner regression, cluster analysis, decision trees, and association analysis (e.g., [28]). NNs have been applied in many areas, from autonomous driving [13] to medical diagnoses [5] to stock market predications [21], for example.

The work with NNs is examined in the remaining part of this paper. The next section deals with its features, affordances, as well as expected challenges and possible reasons for them. The last section summarizes a number of implications for practice, teacher education, and research.

2. Work with neural networks

Features

Artificial neural networks or neural networks (NNs) are nets of interconnected artificial neurons inspired by the biological activity in the brain. These neurons are organized in layers. Each neuron receives one or more weighted inputs, finds their sum, and use it to calculate the output (to this end, a non-linear function is often applied). This output may be an input to other neurons, or the final value of the network exploited (see Fig. 1). After training that fixes all weights through optimization-based machine learning, the final value is typically used for categorization or prediction [26]. Consider NN exploiting data on customers of a mobile phone carrier. This NN might be used to tell the carrier which customers are likely to leave its network on the basis of their age, gender, Internet use, and talking habits. This useful data insight may suggest what special packages could be offered to such customers in order to maintain their loyalty.

Data mining using NNs is usually realized in visual environments, which despite possible complexities regarding their design, interface, and features, simplify the design and application of such computing systems. These environments enable data miners to deal with various NN issues. Basically, these issues are: what NN model to use (e.g., multilayer perception); which NN design to apply (e.g., which input, hidden, and output nodes to use; which non-linear activation function(s) to apply); what parameters of NN training to specify (e.g., what values of leaning rate and momentum to use); and which NN (of possibly several examined and saved) to apply, if any.

Several powerful commercial environments, such as Wolfram Matematica (`https://www.wolfram.com/featureset/data-science/?source=nav`),

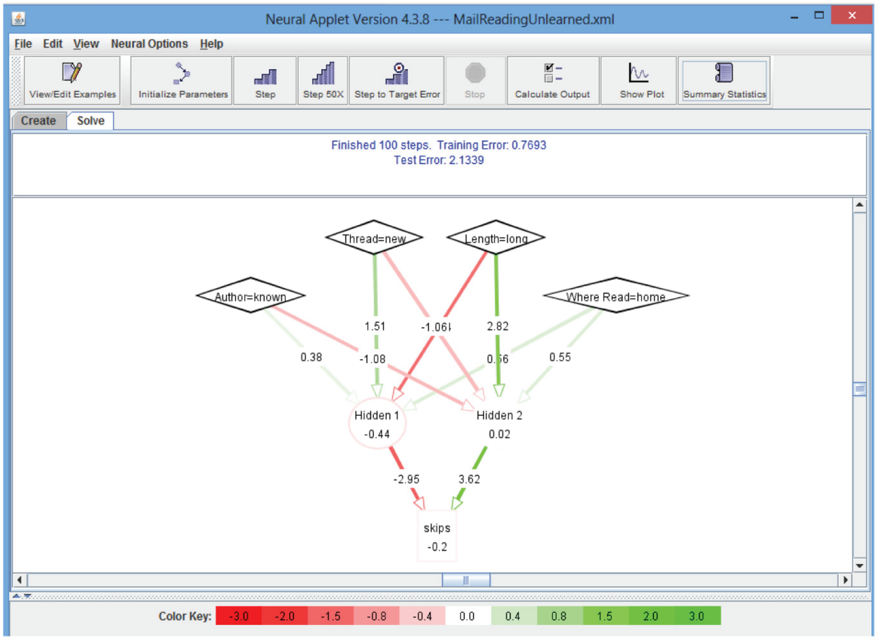


Fig. 1. An example of neural network

Mathlab's Deep Learning Toolbox (`https://www.mathworks.com/products/deep-learning.html`), and StatSoft Statistica (`https://www.statsoft.de/en/statistica/statistica-software`), include sophisticated tools that enable data miners to develop, test, and use their own NNs dealing with previously mentioned issues. On the other hand, there are free (much) simpler environments that support such work to some extent, such as JustNN (`http://www.justnn.com/`), especially a program called Neural Applet (`http://aispace.org/neural/`). This applet is coupled with a number of interesting simple NNs solutions. A screenshot of one of them, with one hidden layer with two neurons (nodes), is presented in Fig. 1. This NN is used to predict the skipping of email by using four binary variables, namely, email author (1-known, 0-unknown), email thread (1-new, 0-old), email length (1-long, 0-short), and e-mail reading place (1-home, 0-other). The available data of 28 records (examples) were randomly split into the training set (with 18 examples) and the validation set (remaining 10 examples) – a somewhat 2/3-1/3 dataset split; other dataset split could be 60%-40% or 70%-30% (these are frequently applied). After training this NN (epochs = 100, learning rate = 0.2, and momentum 0), it attained a satisfactory prediction accuracy of 70% (see Fig. 2).

The prediction accuracy in professional models, based on a richer set of input attributes (predictors), more hidden neurons (with perhaps more hidden layers), and a (considerably) larger dataset, is usually much higher. For the classification of computerized tomographic brain images, for example, the accuracy may be 85% or higher (e.g., [5]). Besides prediction accuracy, which is defined by the proportion

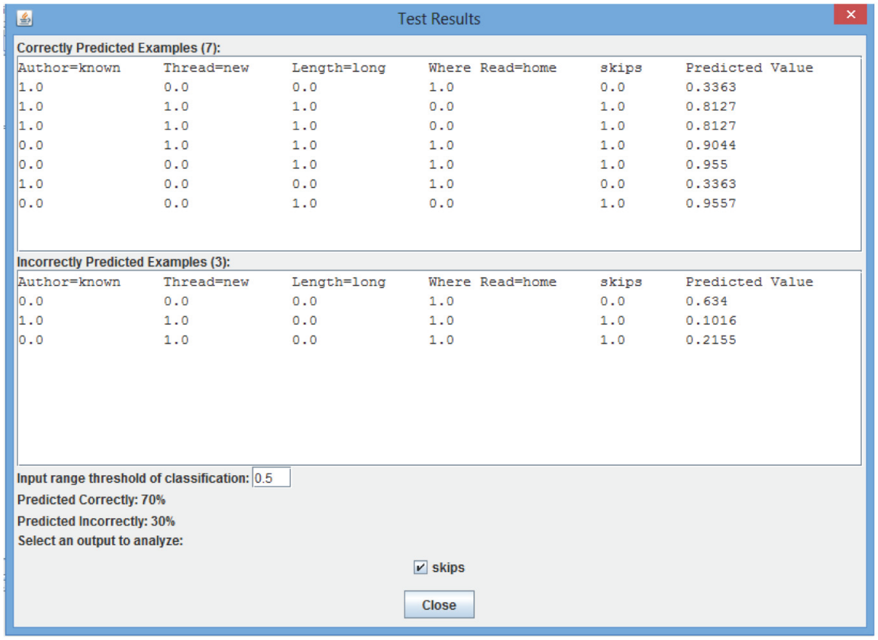


Fig. 2. Predicting the skipping of e-mails and its accuracy

of correct predictions in relation to all predictions made, there are other measures of classification efficiency, such as precision and recall.[1]

What are the main advantages and limitations of using NNs?

According to Shmueli et al. [26], NNs are tolerant to noisy data. They can also capture complex, non-linear relationships between predictors and outcome variables, which is often not possible with other data mining approaches. However, the structure of these relationships is not easily viable due to the black-box character of NN modeling. Other limitations are related to the optimal solution found and the training applied: (1) despite control over learning rate and momentum, NN training may result in weights connected to a local optimum rather than the global one; (2) classifying and predicting would be questionable whenever NNs are over-trained (the so-called overfitting causing a large error on validation data) or trained using unsuitable data (e.g., small data sets or data sets with cases in a limited range). Because NNs do not have a built-in mechanism for predictor selection, they may need to be combined with some dimension reduction approaches (e.g., decision trees) to arrive at key predictors.

[1]The results of binary classifications are usually presented in a 2×2 matrix, whose cells contain the numbers regarding true positives (correct positive classifications), false positives (incorrect positive classifications), true negative, and false negative, which are usually denoted by TP, FP, TN, and FN, respectively. While *precision* stands for TP / (TP + FP), *recall* is defined by TP / (TP + FN). In medical research, frequently applied classification measures are sensitivity and specificity (for these measures, see [23]).

Affordances

Work in an environment for creating, testing, and using NNs may support the understanding of several basic NN concepts, such as node, edge, layer, training sample, activation function, feedforward flow of information, backpropagation scheme, and overfitting. A conceptual clarification of some of these NN concepts (e.g., backpropagation scheme) may support the understanding of general approaches to data modeling (e.g., optimization). Such conceptual clarifications (rather than going into often complex procedural details) in teaching data mining courses for undergraduate business students are proposed by Wu, Mai and Yu [30].

Depending on the complexity of the environment used for creating, testing, and using NNs, this environment may be used to cultivate all or just some of the key stages of the SEMMA cycle. For example, IBM SPSS Statistics with its Neural Networks module (`https://www.ibm.com/analytics/spss-statistics-software`) can be used to cultivate all these key stages, whereas Neural Applet can be used to cultivate only the stages of sampling, modeling, and assessing.

Work with NNs may cultivate data preparing skills because data may need some preparation (e.g., by applying database queries, or organizing data in a different way) before NNs can be built and used. This especially holds true for real case datasets that are often affected by outliers, missing values, and noisy distributions [17]. Sometimes, just saving data in a specific file format may be needed (e.g., csv – comma separated values for Neural Applet). Most of these issues are solely computational, providing an opportunity to learn data science through computation, and vice versa [6].

Because NN learns from training data, it may be unclear what we can learn from data by using NNs. If some of its weights are close to zero, we learn that certain input nodes (input variables) have marginal impact on the output, and, thus, may probably be removed from the NN exploited without reducing its classification accuracy. This kind of analysis may primarily be applied to small NNs. Also, when an NN with good prediction accuracy is found, its application to a set of specific input values can tell us what the output value is likely to be.

It is important to underline that educational studies regarding NNs are not numerous. Most of them deal with improving or evaluating learning or instruction by using NNs that process data relevant to students, teaching, and learning (e.g., [14, 15, 25]). Because of that, educational studies that examine students' work with NNs are quite rare. Exceptions are, for example, the studies of Li et al. [16] and Ugur and Kinaci [28]. The former study examined the NN case teaching approach and found that this approach could increase teaching effectiveness. The latter focused on the use of particular visual tools to teach NN concepts, and reported (among other issues) an improvement of students' success rates. However, these studies did not examine main challenges in the work with NNs, nor did they discuss possible reasons for these challenges.

Challenges

The challenges in using NNs are related to (1) data to use, (2) NN tool to use, and (3) some stages of data mining.

Basically generated by data quality, many challenges may be faced through data preparation, especially when real-case datasets are processed. This is because they are often affected by various data oddities, such as outliers, missing values, and noisy distributions [17]. Before NNs are applied, some of these oddities need to be addressed in satisfactory ways. Another challenge is to use a training dataset of suitable size. For binary classifications, a rule of thumb suggests exploiting a dataset with 100–2,000 training examples [20].

What about challenges regarding which NN tool to use? To use payware, or freeware? When to use them?

Satyanarayana [24] examined various software tools that could be used in an undergraduate data mining course. He recommended using freeware for an introductory course. For an advanced course, he proposed also using a commercial software tool that is applied in industry. Our approach to data modeling with NNs is clearly introductory, and, thus, a free software tool should be used. Which freeware to choose? Obviously, one that supports basic work with NNs by using a simple, user-friendly interface. Such a freeware is, for example, Neural Applet mentioned above.

Challenges are very likely to occur at some stages of data mining. In terms of the SEMMA cycle introduced above, Neural Applet can be used to cultivate the stages of Sampling, Modeling, and Assessing. While Sampling is straightforward, the two other stages may generate challenges. This is because the data miner has to decide on various NN issues, including which NN design to apply (e.g., which input, hidden, and output nodes to use; which activation function(s) to apply); what parameters of NN training to specify (e.g., how many epochs to use, what values of leaning rate and momentum to choose); and which NN (of possibly several examined and saved) to apply, if any.[2] As clear-cut solutions to these issues do not exist, each one may be challenging to address.

Reasons

Some challenges in data preparation may be caused by a limited understanding of basic database concepts, including Boolean logic used in queries, as well as an insufficient experience in building and using simple databases (e.g., [9]). Other challenges, mostly related to data cleaning (e.g., addressing outliers, missing values, and noisy data), are caused by the fact that these is no single way to address each of these data oddities. In a specific case, one solution is usually viewed as more suitable then the other(s). For more details on data cleaning, the first step in data pre-processing, see Chp. 3 in [7], for example.

[2]Initial NN design may be made easier by applying a rule of thumb concerning the number of hidden neurons and the number of hidden layers comprising them. The challenge is that there are several such rules (e.g., [27]). This means that to develop an appropriate NN design that keeps overfitting at bay, a number of trial and error steps needs to be applied.

Considering the use of training datasets of appropriate size, the challenge it generates is primarily caused by the fact that there are several rules of thumb that are not domain independent. This means that while a training dataset with 1,000 examples may be appropriate in one domain, it may be of little use in another, due to its limited representativity of the underlying domain (e.g., [20]).

If challenges in the SEMMA stages are examined in terms of problem structuring skills (applied in [9, 10], for example), it becomes clear that difficulties in making decisions on various NN issues, especially on which NN design to apply, are probably a result of limited skills in problem structuring, as well as the fact that clear-cut solutions (e.g., use one hidden layer with three nodes) do not exists. Furthermore, bearing in mind the importance of data mining context, a failure to use an appropriate set of variables (i.e., a suitable NN design), for example, may be a sign of both limited contextual knowledge and poor problem structuring skills, which are the two inseparable assets in design tasks [22].

3. Implications for practice, teacher education, and research

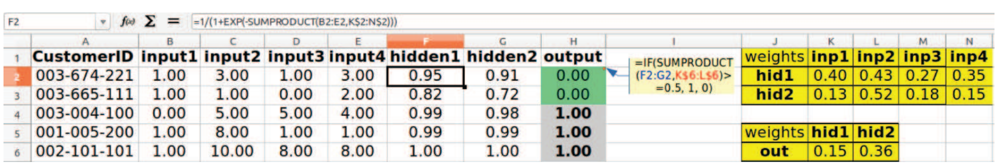
The learning potential of data mining using neural networks was stimulated by the authors experience in teaching a course on decision making with intelligent systems at a private university. There were about twenty forth-year students in one academic year. Most of them developed more or less appropriate NNs (of simple complexity, avoiding overfitting). Although each student chose a problem situation himself/herself, they used fictional, somewhat arbitrary, small datasets, and had difficulties in summarizing actions and decisions supported by their NN models. This showed insufficient familiarity with the context of the situation chosen. Implications for practice, teacher education, and research – based upon this experience, the content of this paper, and other relevant research studies – are summarized in three subsections which follow.

Implications for practice

Introductory work with NNs is accessible to college (business) students. A suitable course may be in mathematics, statistics, or informatics (computing), involving a data science or data mining component. For students with a solid basic knowledge of functions, iterations, and optimization, simple NNs (starting with perception) might be introduced in senior secondary years [12].

To teach basic NN concepts successfully, students should work with NNs by using a simplified visual environment [28]. Furthermore, to help them solve practical problems with NNs, real case-based teaching may be practiced [16]. Although the author of this paper (the instructor) applied this kind of teaching (NNs were developed with Neural Applet), the students' outcomes concerning prescriptive data modeling – i.e., what actions supported by data mining results would be proposed or undertook – were somewhat below expected levels. As mentioned above, each student chose a problem situation himself/herself, but nevertheless, used fictional, and somewhat arbitrary datasets, and had difficulties in summarizing actions and decisions supported by their NN models. These shortcomings were a sign of insufficient familiarity with the context of the situation chosen. This could have improved had the students been supported in developing problem structuring skills and improving knowledge of the context under scrutiny (possibly going back and forth between the two). Because of limited time, the students were not encouraged to experiment with training datasets of different size, different parameters of NN training, or to apply different NN designs (relating to nodes, layers, and activation functions) and compare their validity. Also, to ease data modeling and obtain results for a simplified learning space first, data preparation and data cleaning issues were not covered, only the data packing requested by Neural Applet. These "non-covered" aspects of data modeling with NNs may gradually be included in instruction, defining students' work at different levels of complexity that may follow the SEMMA cycle.[3] A requirement to use real case datasets is often not easy to meet, because many datasets, particularly of business nature, are simply not open to the public.

It may be unclear how to use NN models for prescriptive modeling. Let us return to the NN exploiting data of mobile phone carrier customers introduced in Part 2. Assume that a 4-input attribute NN with good prediction accuracy is found (possibly chosen as the best one of several NNs examined). Assume that this NN tells us who is likely to leave a mobile phone carrier. If we use the data of loyal customers at present, for each one, the NN can tell us whether he or she is likely to leave this carrier in future. If this cannot be attained in the environment used for the work with NNs (like in Neural Applet, for example), the modeler may do this, for example, in a spreadsheet environment as shown in Figure 3 (of course, the spreadsheet model has to use the weights and activation functions from the NN model; instruction should support the development of this spreadsheet solution by providing a template, for example). The carrier can then offer a special package to those whose output values are 1s, and the content of this package may be designed through discovering some patterns present in the input data of those customers (where, for example, the use of interactive charts [10] may prove useful).



F2 =1/(1+EXP(-SUMPRODUCT(B2:E2,K$2:N$2)))

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	CustomerID	input1	input2	input3	input4	hidden1	hidden2	output	=IF(SUMPRODUCT (F2:G2,K$6:L$6)> =0.5, 1, 0)	weights	inp1	inp2	inp3	inp4
2	003-674-221	1.00	3.00	1.00	3.00	0.95	0.91	0.00		hid1	0.40	0.43	0.27	0.35
3	003-665-111	1.00	1.00	0.00	2.00	0.82	0.72	0.00		hid2	0.13	0.52	0.18	0.15
4	003-004-100	0.00	5.00	5.00	4.00	0.99	0.98	1.00						
5	001-005-200	1.00	8.00	1.00	1.00	0.99	0.99	1.00		weights	hid1	hid2		
6	002-101-101	1.00	10.00	8.00	8.00	1.00	1.00	1.00		out	0.15	0.36		

Fig. 3. Supporting prescriptive data modeling: Who is likely to leave mobile phone carrier?

[3]Although the author's teaching experience only dealt with basic activities in constructing and using simple NNs, it is clear that developing (large) NNs of high quality calls for a sophisticated and iterative work concerned with various NN features and parameters. This work is not possible without a wider knowledge of the underlying mathematics and its deeper understanding, which is usually beyond the remit of the college (business) student population that was the target group in this study.

Implications for teacher education

Because of its learning potential, accessibility to students, and relevance to the modern workplace, work with NNs should be covered in teacher education (in mathematics, statistics, or informatics). Apart from the affordances of NNs exemplified through some practical work, this education should cover main expected challenges in using them and possible explanations for these challenges, which would support the (student) teacher in addressing them in appropriate ways in the classroom, referring to the underlying mathematical knowledge in a way accessible to students. Of course, some challenges and explanations may be easier to address than others (e.g., data preparation vs. NN design).

To assist (student) teachers in data mining, teacher education may benefit from carefully structured workshops. During these or other suitable activities, the participants' work may be combined with some Internet-based learning experience [18]. To integrate content (i.e., subject matter), pedagogy, and technology in an appropriate way, teacher education may put into practice the framework called Technological pedagogical content knowledge, bearing in mind that its application may be subject dependent. For example, its application in mathematics (i.e., for mathematics teachers) may assume that technological knowledge integrates with previously integrated pedagogical content knowledge. Its application in statistics may suppose other developmental line: from content knowledge to technological content knowledge to technological pedagogical content knowledge [11]. The developmental line in informatics (computing) may differ from these two lines because informatics (computer science) teachers usually have a solid knowledge of technology. Of course, the integration process of these three forms of knowledge may differ from teacher to teacher.

Implications for research

On the basis of the previous discussion, several directions for further research may be suggested. Some of them are given below, but their order does not reflect their relative importance.

Data scientists spend most of their time collecting, cleaning, and organizing data (e.g., [4]). There are many challenges concerning these activities, and a number of sources that can generate these challenges. Research may thus focus on developing and testing a computer adviser (e.g., an expert system; cf. [2]) that supports students in developing appropriate datasets for data mining considering their content and size. Of course, such an adviser may be developed for other perplexing topics in data mining, where solutions are based upon heuristics or rules of thumb rather than exact rules. To make these topics more accessible to miners, the adviser may explain and connect pieces of underlying knowledge from mathematics, statistics, and computing (computer science).

Students should be encouraged and supported to use data mining not only to understand the area under scrutiny, but also to make suggestions about how to change or improve it. However, without solid domain knowledge, the majority of students are likely to fail in making (model-grounded and context-grounded) suggestions and proposals of actions to take (e.g., [8]). Although students should always base modeling activities on contexts familiar to them (personal communication with Henry Pollak, July 26, 2016, ICME-13, Hamburg, Germany), many of them may not be (that) familiar even with contexts they have chosen themselves. Because contextual knowledge and problem structuring skills probably influence each other (extrapolated from [22]), research could examine ways that support data miners in developing problem structuring skills while simultaneously improving knowledge of the context under scrutiny.

An important component of data mining courses is to cultivate communication skills, which are consistently ranked by employers among the most important skills (e.g., [30]). This means that instruction should support students in developing these skills. Research could thus study basic communications skills used by successful data miners and use the outcomes to inform and support instruction in improving students' communication ability.

A recent review of research on business intelligence & analytics (BI & A) education found that, among other research gaps, research studies on BI & A pedagogical and learning issues are mostly exploratory or based upon case studies. Furthermore, "theoretical frameworks with empirical or mixed methodologies for assessing the students' learning effectiveness on the BI & A knowledge and skills are lacking." ([29], p. 4). Research should thus begin to combine theoretical and empirical aspects of data mining. The author hopes that the research presented in this paper may contribute to this research direction.

Acknowledgement. The author wishes to thank the anonymous referee whose suggestions helped to improve this contribution. It resulted from his research funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-68/2020-14/200018). The author dedicates the contribution to his son Aleksandar.

REFERENCES

[1] Asia Pacific Economic Cooperation (APEC) Agency (2017), *Data science and analytics skills shortage: Equipping the APEC workforce with the competencies demanded by employers*, Singapore, The Author. `https://www.apec.org/Publications/2017/11/Data-Science-and-Analytics-Skills-Shortage`

[2] S. Brijendra, P. Dhanavanthan, T. Mani, *SESYS: Statistical expert systems for information management*, IETE Technical Review, **21** (2) (2004), 137–148. `https://doi.org/10.1080/02564602.2004.11417140`

[3] L. Cao, *Data science: A comprehensive overview*, ACM Computing Surveys, **50** (3) (2017), 43:1–42. `https://doi.org/10.1145/3076253`

[4] Crowdflower, *Data science report*, The Author. `https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf`

[5] X. W. Gao, R. Hui, Z. Tian, *Classification of CT brain images based on deep learning networks*, Computer Methods and Programs in Biomedicine, **138** (2017), 49–56. `https://doi.org/10.1016/j.cmpb.2016.10.007`

[6] R. Gould, A. Bargagliotti, T. Johnson, *An analysis of secondary teachers' reasoning with participatory sensing data*, Statistics Education Research Journal, **16** (2) (2017), 305–334. `https://iase-web.org/documents/SERJ/SERJ16(2)_Gould.pdf?1512143292`

[7] J. Han, M. Kamber, J. Pei, *Data mining: Concepts and techniques* (3rd ed), Amsterdam: Elsevier, 2012. https://doi.org/10.1016/C2009-0-61819-5

[8] D. Kadijevich, *Examining errors in simple spreadsheet modeling from different research perspectives*, Journal of Educational Computing Research, **47** (2) (2012), 137–153. https://doi.org/10.2190/EC.47.2.b

[9] D. M. Kadijevich, *Data modeling with dashboards: opportunities and challenges*, In: J. Engel (Ed.), *Promoting understanding of statistics about society. Proceedings of the roundtable conference of the International Association of Statistics Education (IASE), July 2016*, Berlin, Germany, ISI/IASE, the Haag, the Netherlands, 2016. https://iase-web.org/documents/papers/rt2016/Kadijevich.pdf?1482482774

[10] D. M. Kadijevich, *Data modeling using interactive charts*, The Teaching of Mathematics, **21** (2) (2018), 55-72. http://www.teaching.math.rs/vol/tm2121.pdf

[11] D. M. Kadijevich, S. R. Madden, *Comparing approaches for developing TPCK*, In: Khine, M. S. (Ed.), *New directions in technological pedagogical content knowledge research: Multiple perspectives* (pp. 125-146). Charlotte, NC: Information Age Publishing, 2015.

[12] D. M. Kadijevich, M. Stephens, *Modern statistical literacy, data science, dashboards, and automated analytics and its applications*, The Teaching of Mathematics, **23** (1) (2020), 71–80. http://www.teaching.math.rs/vol/tm2318.pdf

[13] J. Kocić, N. Jovičić, V. Drndarević, *An end-to-end deep neural network for autonomous driving designed for embedded automotive platforms*, Sensors, **19** (9) (2019), 2064. https://doi.org/10.3390/s19092064

[14] U. Kose, A. Arslan, *Optimization of self-learning in computer engineering courses: An intelligent software system supported by artificial neural network and vortex optimization algorithm*, Computer Applications in Engineering Education, **25** (1) (2017), 142–156. https://doi.org/10.1002/cae.21787

[15] R. L. Lamb, D. B. Vallett, T. Akmal, K. Baldwin, *A computational modeling of student cognitive processes in science education*, Computers & Education, **79** (C) (2014), 116–125. https://doi.org/10.1016/j.compedu.2014.07.014

[16] J. Li, H. Zhang, Y. Zhou, Y. Bai, *RBF neural network case teaching research*, In: H. Tan, & M. Zhou (Eds.), *Advances in information technology and education* (pp. 351-355), Communications in Computer and Information Science, vol. 201. Berlin: Springer, 2011. https://doi.org/10.1007/978-3-642-22418-8_49

[17] F. Maiorana, *A teaching experience on a data mining module*, In: *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Wroclaw, 2012 (pp. 871–874). New York: IEEE, 2012.

[18] A. McNamara, M. Hansen, *Teaching data science to teenagers*, In: Makar, K., de Sousa, B., & Gould, R. (Eds.), *Proceedings of the Ninth international conference on teaching statistics* (CD-ROM), Voorburg, The Netherlands: International Statistical Institute, 2014. https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_9C2_MCNAMARA.pdf?1405041770

[19] D. L. Olson, D. Delen, *Data mining process*, In: Olson, D. L., & Delen, D., *Advanced data mining techniques* (pp. 9–35). Berlin: Springer, 2008. https://doi.org/10.1007/978-3-540-76917-0_2

[20] T. Oyedare, J. J. Park, *Estimating the required training dataset size for transmitter classification using deep learning*, In: *2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Newark, NJ, USA, November 11–14 (pp. 1–10). New York: IEEE, 2019. https://doi.org/10.1109/DySPAN.2019.8935823

[21] X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, *An innovative neural network approach for stock market prediction*, The Journal of Supercomputing **76**, (1) (2020), 2098–2118. https://doi.org/10.1007/s11227-017-2228-y

[22] J. Restrepo, H. Christiaans, *Problem structuring and information access in design*, Journal of Design Research, **4** (2) (2004). https://doi.org/10.1504/JDR.2004.009842

[23] T. Saito, M. Rehmsmeier, *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets*, PLoS ONE, **10** (3) (2015), e0118432. https://doi.org/10.1371/journal.pone.0118432

[24] A. Satyanarayana, *Software tools for teaching undergraduate data mining course*, Paper presented at American Society of Engineering Education Mid-Atlantic Fall Conference, October 2013, Washington, DC. New York: New York City College of Technology, 2013.

[25] Z. Sevarac, V. Devedzic, J. Jovanovic, *Adaptive neuro-fuzzy pedagogical recommender*, Expert Systems with Applications, **39** (10) (2012), 9797–9806. `https://doi.org/10.1016/j.eswa.2012.02.174`

[26] G. Shmueli, N. R. Patel, P. C. Bruce, *Data mining for business intelligence*, Hoboken, NJ: John Wiley & Sons, 2007.

[27] D. Stathakis, *How many hidden layers and nodes?* International Journal of Remote Sensing, **30** (8) (2009), 2133–2147. `https://doi.org/10.1080/01431160802549278`

[28] A. Ugur, A. C. Kinaci, *A web-based tool for teaching neural network concepts*, Computer Applications in Engineering Education, **18** (3) (2010), 449–457. `https://doi.org/10.1002/cae.20184`

[29] Y. Wang, *Business intelligence and analytics education: Hermeneutic literature review and future directions in IS education*, In: *Proceeding of Twenty-First Americas Conference on Information Systems (AMCIS)*, Puerto Rico, 2015. Atlanta, GA: AIS.

[30] C. Wu, F. Mai, Y. Yu, *Teaching data mining to business undergraduate students using R*, Business Education Innovation Journal, **7** (2) (2015), 64–73. `http://www.beijournal.com/images/V7N2_final.pdf`

D.M.K.: Institute for Educational Research, Dobrinjska 11/III, Belgrade, Serbia
E-mail: `djkadijevic@ipi.ac.rs`