International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 3 Issue 10, October – 2019, Pages: 1-8

Machine Learning Application to Predict The Quality of Watermelon Using JustNN

Ibrahim M. Nasser
Gaza, Palestine
Azhar.ibrahimn@gmail.com

***Abstract*:** In this paper, a predictive artificial neural network (ANN) model was developed and validated for the purpose of prediction whether a watermelon is good or bad, the model was developed using JUSTNN software environment. Prediction is done based on some watermelon attributes that are chosen to be input data to the ANN. Attributes like color, density, sugar rate, and some others. The model went through multiple learning-validation cycles until the error is zero, so the model is 100% percent accurate for the purpose of prediction whether good or bad the watermelon is.

Keywords: Machine Learning; Deep Learning; Predictive Analysis; Artificial Neural Network; JUSTNN

1. INTRODUCTION

Machine Learning (ML) is a use of artificial intelligence (AI) which provides systems the ability to automatically learn and improve from practice without being obviously programmed. ML emphases on the change of computer programs that can access data and use it to learn for themselves. The method of learning starts with data to look for patterns in that data and create enhanced choices in the future based on the data provided. The key purpose is to allow the computers learn automatically lacking human interference or help and modify actions so.

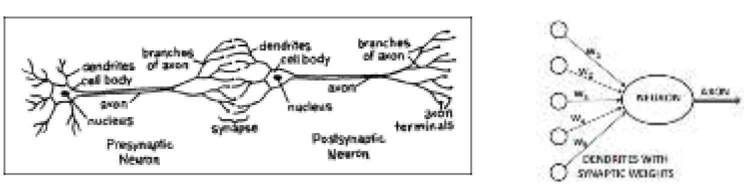
ML algorithms can be categorized as supervised or unsupervised. Supervised algorithms can apply what has been learned in the past to new data examples to predict future events. Starting from the study of a known training dataset, the learning algorithm produces a function to make expectations about the output values. The system is able to provide outputs for any new input after adequate training. The learning algorithm can also compare its output with the correct one, and find errors so it can adapt the model consequently.

On the other hand, unsupervised algorithms are used when the data used to train is neither classified nor labeled. Unsupervised learning studies how systems can assume a function to describe a hidden structure from unlabeled data. The system does not discover the right output, but it explores the data and can draw interpretations from datasets to describe hidden structures from unlabeled data [1].

Deep Learning (DL) is a subfield of ML concerned with algorithms inspired by the construction and role of the brain called Artificial Neural Networks (ANN) [2].

2. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANNs) are common ML methods that stimulate the process of learning in biological organism. The human nervous system contains cells, which named neurons. The neurons are connected to each other using axons and dendrites, and the connection regions between axons and dendrites are referred to as synapses. These connections are showed in Figure 1 (a). The strengths of synaptic links often change in response to external stimuli. This change is how learning is done in living creatures. This biological mechanism is stimulated in ANNs, which contain computation units named neurons. The computational units are associated to one another through weights, which function the same as the synaptic strength connections in biological creatures. Each input to a neuron is scaled with a weight, which affects the function computed at that unit [3]. This architecture is shown in Figure 1 (b).



(a) Biological Neural Network (b) Artificial Neural Network

Figure 1: BNN& ANN. The image (a) from [4]. The image (b) from [3].

In training, first, weights are assigned with random numbers, then the validity of the ANN is examined and errors calculated. Next the weights are adjusted based on validation results. This process keeps repeating until the ANN got an acceptable validation error rate [5]. Neurons often are grouped into three layers. First, The input layer, which responsible for receiving input from the user, second, the output Layer, which sends data to the user, finally, the hidden layer(s) between the input and output layer. General structure is shown in Figure 2.

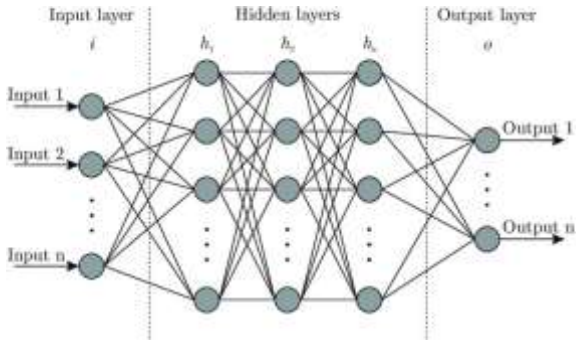


Figure 2: ANN Architecture

In validation, validation data must be different from the training data to correctly validate the ANN model [6]. In this application, I used about 60% of the data for model training, and about 40% for validation.

3. JUSTNN

JUSTNN is a free software application for Microsoft Windows. It makes the design of neural networks easy. It permits the user to build multilayer neural networks. JUSTNN can perform learning from training data and can validate while learning. It can be queried from a file or interactively. JUSTNN can produce *TXT* or *CSV* output and results files [8].

4. LITERATURE REVIEW

Until recently, many ANNs models were created that serving predicting and classification purposes. Abu Naser, S., et al. built a model that able to predict learners performance with 92% accuracy [9] and many other models [20- 34].
Al-Shawwa, M., et al. developed many ANNs models e.g. to predict birth weight, effect of oxygen consumption of Thylakoid Membranes (Chloroplasts) from Spinach after Inhibition, and temperature and humidity in the surrounding environment [10-12].
Nasser, I. M., et al. worked also in deep learning applications, and they have done a bunch of useful models e.g. Diagnose Autism Spectrum Disorder, Lung Cancer Detection, Predicting Tumor Category, and some other applications [13- 19].

5. THE DATASET

The dataset used is named *xigua3.0 (Watermelons)*, which I got from *Kaggle*, and its author is *benjaminwang [7]*.
Dataset Description is shown in table 1.

Table 1: Original Dataset Description

Column	Values
Color	Green, black, pale, light white
Pedicle	Curled, slightly sloppy, stiff
Knock	Turbid, dull, crisp
Texture	Clear, slightly paste, fuzzy
Umbilical	Sunken, concave, slightly concave, flat, sag
Touch	Hard slip, soft sticky, soft
Density	Positive real numbers
Sugar rate	Positive real numbers
Good melon (class)	Yes, No

6. DATASET PREPROCESSING

The original data was in Chinese language, so I translated it to English, and then I did some preprocessing to the values so it becomes more suitable to the ANN model.

All text data were transformed to numerical and Boolean data, the new dataset description is shown in table 2.

Table 2: Dataset Preprocessing

Column	Values
Color	[1- 5]
Pedicle	[1- 3]
Knock	[1- 3]
Texture	[1 -3]
Umbilical	[1 -5]
Touch	[1 -3]
Density	Positive real numbers
Sugar rate	Positive real numbers
Good melon (class)	True, False

7. METHODOLGY

I entered the data file as input to the model, then I chose the class to be predicted, then the environment pick randomly instances for validation depends on the percentage I asked for. There were some validation instances that out of validation range, I changed all of them to training and chose other instances.

Then I did the settings to the network so it starts learning and validation, I keep changing settings until the model got an acceptable error rate.

Training and validation instances are shown in figure 3.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 3 Issue 10, October – 2019, Pages: 1-8

	color	pedicle	knock	texture	umbilical	touch	density	sugar rate	good melon
#0	1	1	1	1	1	1	0.6970	0.4600	true
#1	2	1	2	1	1	1	0.7740	0.3760	true
#2	2	1	1	1	1	1	0.6340	0.2640	true
#3	1	1	2	1	1	1	0.6080	0.3180	true
#4	3	1	1	1	2	1	0.5560	0.2150	true
#5	1	2	1	1	3	2	0.4030	0.2370	true
#6	2	2	1	2	3	2	0.4810	0.1490	true
#7	2	2	1	1	3	1	0.4370	0.2110	true
#8	2	2	2	2	3	1	0.6660	0.0910	false
#9	1	3	3	1	4	3	0.2430	0.2670	false
#10	4	3	3	3	4	1	0.2450	0.0570	false
#11	3	1	1	3	4	3	0.3430	0.0990	false
#12	1	2	1	2	5	1	0.6390	0.1610	false
#13	4	2	2	2	1	1	0.6570	0.1980	false
#14	2	2	1	1	3	2	0.3600	0.3700	false
#15	5	1	1	3	4	1	0.5930	0.0420	false
#16	1	1	2	2	3	1	0.7190	0.1030	false

Figure 3: The data grid (greens are validation instances, blues are for training)

8. RESULTS

After 1425201 training-validation cycles, the ANN model was successfully created with about 0 error and 100% accuracy as shown in figure 4.

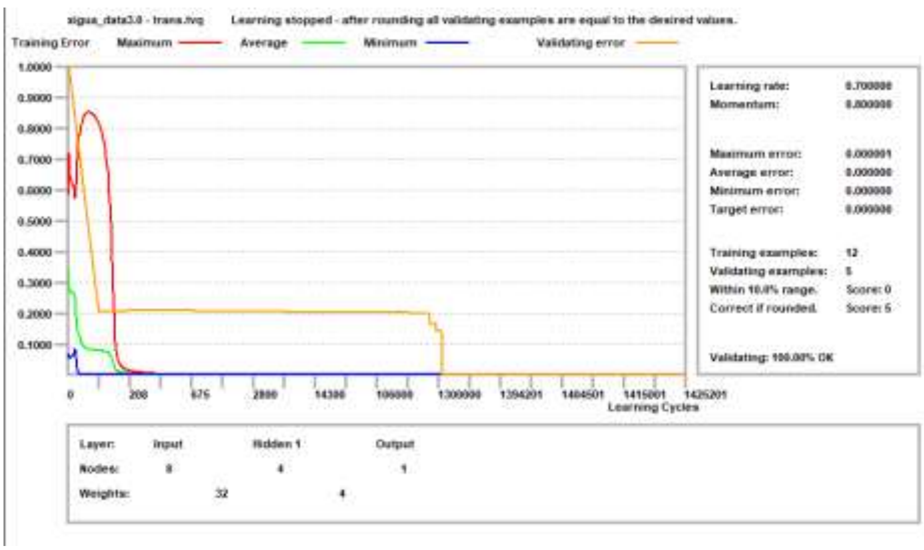


Figure 4: Validation errors chart

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 3 Issue 10, October – 2019, Pages: 1-8

Moreover, the model discovered that the attribute *knock* is the most important one in the input data, more details are shown in figure 5.

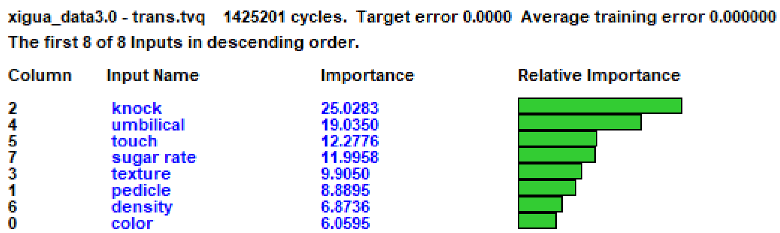


Figure 5: Attributes importance

The controls I set are shown in figure 6.

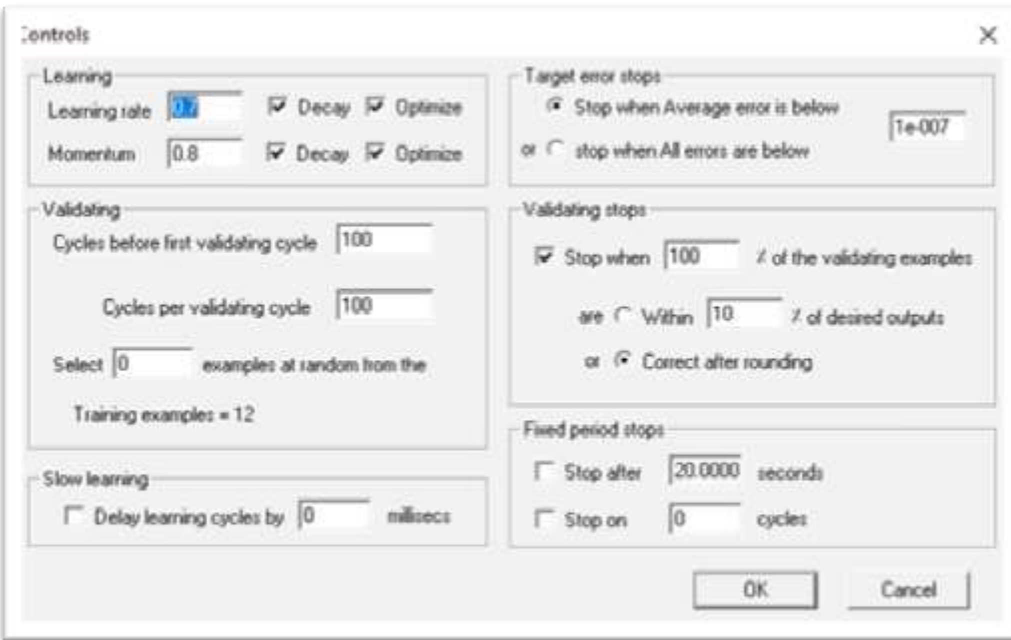


Figure 6: ANN model controls

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 3 Issue 10, October – 2019, Pages: 1-8

Finally, the ANN model that resulted is shown in figure 7.

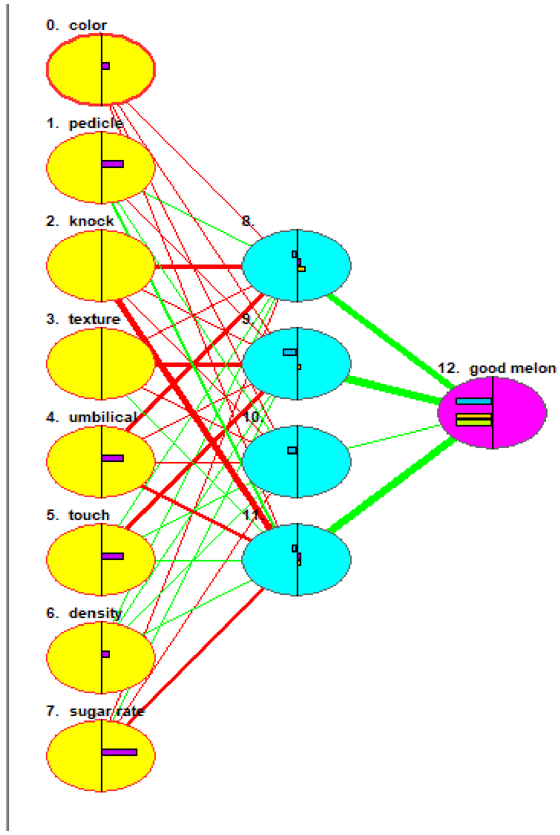


Figure7: The ANN model

9. LIMITATIONS

The dataset is not as large as it should be; it contains just 17 instances, which I split them to 12 for training and 5 for validation.

10. CONCLUSION

In this research paper, I developed an Artificial Neural Network that able to predict which the watermelon is good or bad based on some data about that watermelon. Validation showed that the ANN model is 100% accurate to do that job with zero error rate. The network is developed, trained and validated using JUSTNN Software and a dataset I got form KAGGLE.

REFERENCES

1. "What is Machine Learning? A definition," 7 March 2017. [Online]. Available: https://expertsystem.com/machine-learning-definition/. [Accessed 12 October 2019].
2. "Deep Learning & Artificial Neural Networks," 16 August 2019. [Online]. Available: https://machinelearningmastery.com/what-is-deep-learning/. [Accessed October 2019].
3. C. C. Aggarwal, *Neural Networks and Deep Learning. A Textbook*, Springer, 2018, pp. 1 - 2.
4. https://science.education.nih.gov/supplements/webversions/BrainAddiction/guide/lesson2-1.html
5. P. Khanale and S. Chitnis, “Handwritten Devanagari Character Recognition using Artificial Neural Network”, Journal of Artificial Intelligence, vol. 4, no. 1, (2011).
6. H. Martin and D. Howard, “Neural Network Design”, 2nd Edition, Martin Hagan (2014).
7. "xigua3.0 (Watermelons)," 04 05 2018. [Online]. Available: https://www.kaggle.com/benjaminwang/xigua30. [Accessed October 2019].

8. "Soft112," [Online]. Available: https://justnn.soft112.com. [Accessed October 2019].
9. Abu Naser, S. S. (2012). "Predicting learners performance using artificial neural networks in linear programming intelligent tutoring system." International Journal of Artificial Intelligence & Applications 3(2): 65.
10. Al-Shawwa, M. and S. S. Abu-Naser (2019). "Predicting Effect of Oxygen Consumption of Thylakoid Membranes (Chloroplasts) from Spinach after Inhibition Using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 3(2): 15- 20.
11. Al-Shawwa, M., et al. (2018). "Predicting Temperature and Humidity in the Surrounding Environment Using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 2(9): 1-6.
12. Al-Shawwa, M. and S. S. Abu-Naser (2019). "Predicting Birth Weight Using Artificial Neural Network." International Journal of Academic Health and Medical Research (IJAHMR) 3(1): 9-14.
13. Nasser, Ibrahim M., and Samy S. Abu-Naser. "Predicting Tumor Category Using Artificial Neural Networks." (2019).
14. Nasser, Ibrahim M., Mohammed O. Al-Shawwa, and Samy S. Abu-Naser. "Artificial Neural Network for Diagnose Autism Spectrum Disorder." (2019).
15. Nasser, Ibrahim M., and Samy S. Abu-Naser. "Lung Cancer Detection Using Artificial Neural Network." International Journal of Engineering and Information Systems (IJEAIS) 3.3 (2019): 17-23.
16. Nasser, Ibrahim M., and Samy S. Abu-Naser. "Artificial Neural Network for Predicting Animals Category." (2019).
17. Nasser, Ibrahim M., Mohammed O. Al-Shawwa, and Samy S. Abu-Naser. "A Proposed Artificial Neural Network for Predicting Movies Rates Category." (2019).
18. Nasser, Ibrahim M., Mohammed O. Al-Shawwa, and Samy S. Abu-Naser. "Developing Artificial Neural Network for Predicting Mobile Phone Price Range." (2019).
19. Nasser, Ibrahim M., and Samy S. Abu-Naser. "Predicting Books' Overall Rating Using Artificial Neural Network." (2019).
20. Afana, M., et al. (2018). "Artificial Neural Network for Forecasting Car Mileage per Gallon in the City." International Journal of Advanced Science and Technology 124:51-59
21. Alghoul, A., et al. (2018). "Email Classification Using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 2(11): 8-14.
22. Alkronz, E. S., et al. (2019). "Prediction of Whether Mushroom is Edible or Poisonous Using Back- propagation Neural Network." International Journal of Academic and Applied Research (IJAAR) 3(2): 1-8.
23. Al-Massri, R., et al. (2018). "Classification Prediction of SBRCTs Cancers Using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 2(11): 1-7.
24. Al-Mubayyed, O. M., et al. (2019). "Predicting Overall Car Performance Using Artificial Neural Network." International Journal of Academic and Applied Research (IJAAR) 3(1): 1-5.
25. Ashqar, B. A., et al. (2019). "Plant Seedlings Classification Using Deep Learning." International Journal of Academic Information Systems Research (IJAISR) 3(1): 7-14.
26. El_Jerjawi, N. S. and S. S. Abu-Naser (2018). "Diabetes Prediction Using Artificial Neural Network." International Journal of Advanced Science and Technology 121: 55-64.
27. El-Khatib, M. J., et al. (2019). "Glass Classification Using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 3(2): 25- 31.
28. Elzamly, A., et al. (2017). "Predicting Critical Cloud Computing Security Issues using Artificial Neural Network (ANNs) Algorithms in Banking Organizations." International Journal of Information Technology and Electrical Engineering 6(2): 40-45.
29. Jamala, M. N. and S. S. Abu-Naser (2018). "Predicting MPG for Automobile Using Artificial Neural Network Analysis." International Journal of Academic Information Systems Research (IJAISR) 2(10): 5-21.
30. Kashf, D. W. A., et al. (2018). "Predicting DNA Lung Cancer using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 2(10): 6-13.
31. Metwally, N. F., et al. (2018). "Diagnosis of Hepatitis Virus Using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 2(11): 1-7.

32. Sadek, R. M., et al. (2019). "Parkinson‘s Disease Prediction Using Artificial Neural Network." International Journal of Academic Health and Medical Research (IJAHMR) 3(1): 1-8.
33. Salah, M., et al. (2018). "Predicting Medical Expenses Using Artificial Neural Network." International Journal of Engineering and Information Systems (IJEAIS) 2(20): 11-17.
34. Zaqout, I., et al. (2015). "Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology." International Journal of Hybrid Information Technology 8(2): 221-228.

Authors

Ibrahim M. Nasser

Ibrahim joined the Department of Software Engineering at Al-azhar University - Gaza in 2014. He graduated recently and currently he is an independent Researcher, he has done research works with Professor Sami S. Abu Naser, He is interested in Artificial Intelligence Research, in particular deep learning techniques.