International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 9, September – 2020, Pages: 62-67

Artificial Neural Network for Predicting Workplace Absenteeism

Raghad Adnan Abu Hassanein, Saja Ahmed Al-Qassas, Fatima Naji Abu Tir, Samy S. Abu-Naser

Department of Information Technology,
Faculty of Engineering and Information Technology,
Al-Azhar University, Gaza, Palestine

***Abstract**: Associations can grow, succeed, and sustain if their employees are committed. The main assets of an association are those employees who are giving it a required number of hours per month, in other words, those employees who are punctual towards their attendance. Absenteeism from work is a multibillion-dollar problem, and it costs money and decreases revenue. At the time of hiring an employee, Associations do not have an objective mechanism to predict whether an employee will be punctual towards attendance or will be habitually absent. For some Associations, it can be very difficult to deal with those employees who are not punctual, as firing may be either not possible or it may have a huge cost to the association. In this paper, we propose An Artificial Neural Networks algorithm that can predict the behavior of employees towards punctuality at workplace. The efficacy of the proposed method is tested with traditional machine learning techniques, and the results indicate 99.00% accuracy. The proposed model will provide a useful mechanism to associations that are interested to know the behavior of employees at the time of hiring and can reduce the cost of paying to inefficient or habitually absent employees.*

Keywords: Artificial Neural Network, Prediction, Workplace, Absenteeism

Introduction

The growth and success of an Association depend on its employees. Therefore, it is really important that employees are punctual towards their attendance and work for the number of hours defined by the employer Associations. Generally, the number of working hours in an Association is eight hours per day or 240 hours per month. Associations prefer that employees are present for the maximum number of hours. However, there can be unavoidable circumstances, and hence, different types of leaves are granted in most of the cases. In different countries, employees are allowed to take a certain amount of leaves per month, and employers expect employees to use their due right. But those who are habitually absent are the real reason why Association's productivity or revenue decreases.

Absenteeism at work can be defined as a habitual pattern of absence from a duty or obligation. Generally, absenteeism is assumed as a major indicator of poor performance. Absenteeism happens when employees are habitually late or engage in activities that are not directly or indirectly related to their work, e.g., long coffee breaks, overextended breaks, excessive personal times, internet time, and unnecessary socialization. At work places, different strategies are used to enforce effective time utilization [1]. For instance, some Associations enforce a check-in and check-out time and may deploy different software or biometric devices to detect absenteeism at work. The behavior of employees towards absenteeism cannot be straightforward to detect and can come in numerous shapes and several stages of severity [2–4]. For instance, an employee may be sitting in his/her office for eight hours a day but may be involved in lengthy phone calls, social networking sites, games, etc. It is very important for the Association to predict the behavior of employees towards punctuality at work at an early stage, such as before calling them for interview for a vacant position.

This can avoid the cost of conducting interview, hiring process, and even the hassle of dealing with people who are habitually absent and are affecting the Associational work environment.

Neural Networks (NNs) are around for many years and have been used extensively in solving different problems [5]. In the last few years, Neural Network, particularly Deep Neural Networks (DNNs), is becoming extremely popular [6, 7] and is achieving better performance compared to traditional machine learning algorithms (such as logistic regression, Decision Tree, and SVM). In this paper, we present Neural Network and Deep Neural Network that can predict the behavior of employees towards punctuality in attendance. Associations can deploy these models to predict the behavior of employees towards punctuality at workplace and can make appropriate decisions about the selection of employees at the time of making new intakes.

The main contribution of this paper is to analyze different parameters which can potentially contribute to absenteeism and then develop a model based on JNN to most accurately predict the absenteeism before employees are actually hired.

2. Literature Review

Artificial Neural Networks have been used many fields. In Education such as: Predicting Student Performance in the Faculty of Engineering and Information Technology using ANN, Prediction of the Academic Warning of Students in the Faculty of Engineering and Information Technology in Al-Azhar University-Gaza using ANN, Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach[6].

In the field of Health such as: Parkinson's Disease Prediction, Classification Prediction of SBRCTs Cancers Using ANN, Predicting Medical Expenses Using ANN, Predicting Antibiotic Susceptibility Using Artificial Neural Network, Predicting Liver Patients using Artificial Neural Network, Blood Donation Prediction using Artificial Neural Network, Predicting DNA Lung Cancer using Artificial Neural Network, Diagnosis of Hepatitis Virus Using Artificial Neural Network, COVID-19 Detection using Artificial Intelligence[7].

In the field of Agriculture: Plant Seedlings Classification Using Deep Learning , Prediction of Whether Mushroom is Edible or Poisonous Using Back-propagation Neural Network, Analyzing Types of Cherry Using Deep Learning, Banana Classification Using Deep Learning, Mango Classification Using Deep Learning, Type of Grapefruit Classification Using Deep Learning, Grape Type Classification Using Deep Learning, Classifying Nuts Types Using Convolutional Neural Network, Potato Classification Using Deep Learning, Age and Gender Prediction and Validation Through Single User Images Using CNN[8].

In other fields such as : Predicting Software Analysis Process Risks Using Linear Stepwise Discriminant Analysis: Statistical Methods, Predicting Overall Car Performance Using Artificial Neural Network, Glass Classification Using Artificial Neural Network, Tic-Tac-Toe Learning Using Artificial Neural Networks, Energy Efficiency Predicting using Artificial Neural Network, Predicting Titanic Survivors using Artificial Neural Network, Classification of Software Risks with Discriminant Analysis Techniques in Software planning Development Process, Handwritten Signature Verification using Deep Learning, Email Classification Using Artificial Neural Network, Predicting Temperature and Humidity in the Surrounding Environment Using Artificial Neural Network, English Alphabet Prediction Using Artificial Neural Networks[9].

In the Absenteeism field, there are a few studies such as in [9-10] proposed a model which can predict the number of hours an employee is absent from work. Four different categories of absenteeism are used. Similarly, another research study

In another research study by the author in [11], a pilot's absenteeism is predicted in an airline company. In the airline industry, crew costs are the second most important cost after fuel costs, and pilots are the most important airline crew. For airline companies, having a system that can predict pilot absenteeism can help to manage the operations. They are using the Decision Tree algorithm in building a decision support system to predict the number of hours a pilot will be absent and to make necessary arrangements to deal with the situation. This system is dealing only with pilot absenteeism and can be used in an airline industry.

In the literature, there are a number of research studies which are using machine learning and other data mining techniques to understand the hidden patterns in the data. These machine learning and data mining algorithms are extensively used as classification models to predict different patterns in dataset. For example, in [12], a Decision Tree is used as a data mining technique to predict the attendance pattern of employees. In that research study, a private company's data are used as a case study to test classification algorithms. Similarly, in [12], machine learning models are used to predict employee turnover. Also, in another study, data mining techniques are used to predict employee turnover [13]. In [14], machine learning algorithms (Random Forest) are used to predict dropout in high school students. In another study, the machine learning algorithm is used to identify students at risk of adverse academic outcomes [15] and also, the data mining technique was used to predict secondary school student performance is presented in [16]. Traditional machine learning algorithms known as Decision Tree, Gradient Boosted Tree, Random Forest, and Tree Ensemble were used for absenteeism in [17]. The paper reports an accuracy rate of 82% by the Gradient Boosted Tree, while Tree Ensemble performed the lowest in terms of the accuracy rate of 97%. Authors in [18] have demonstrated the resource management and scheduling based on Stochastic-Petri Net Modeling and optimization for patients to make a sustainable healthcare system, which takes into account the absenteeism factor of medical staff. The study demonstrates that when the absenteeism factor of medical staff is taken into account, the performance of the healthcare system is improved significantly in terms of reduced waiting time for patients and improved operational sustainability. DNNs in the context of predicting image privacy have been studied in [19,20].

All previous studies have applied machine learning and deep learning techniques to model diverse problems. They are based on the understanding that the pattern of absenteeism at workplace and proposed different solutions to reduce the absenteeism rate. However, they did they did not consider prediction of absenteeism at early stage as proposed in the current study. Clearly, there is a research gap for modeling and predicting absenteeism behavior of employees at workplace at the early stage of their hiring. The models presented in this paper are developed as general models and can predict employee's behavior at the time of hiring, whether the employee will be punctual or tend to be more frequently absent in future.

3. Methodology

3.1 Dataset

In this research study, workplace absenteeism data are taken from the UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)[21].
The dataset contains a total of 20 different features and 740 samples. These data samples reflect the behavior of employees towards punctuality at a courier company in Brazil.

These features are:
1 Reason for absence,
2 Month of absence,
3 Day of the week,
4 Seasons,
5 Transportation expense,
6 Distance from residence to work, Service time,
7 Age,
8 Work load average/day,
9 Hit target,
10 Disciplinary failure, Education, Son,
11 Social drinker,
12 Social smoker,
13 Pet,
14 Weight,
15 Height,
16 Body mass index, and
17 Absenteeism category.

3.2. Data Analysis.
We have designed the absenteeism category as *moderate* where the number of hours an employee is absent for 0–5 hours per month, and the other class is *excessive* when the number of hours an employee is absent for more than 5 hours. This extra number of hours of absence is calculated after the leaves that are allowed for employees as per Association's policy (i.e., one or two days per month and paternity/maternity leave). A relaxation of 5 hours is given in the model, only to differentiate between employees who happen to have a problem, compared to the employee who is habitually absent. But after 5 hours, Associations indicate their concerns about the number of hours employees are absent from work. The dataset is small, containing only 740 instances. It is not a problem for traditional machine learning algorithms such as SVM [11] or Logistic Regression [12], but Neural Network and particularly Networks are data hungry. We demonstrate the results of Just Neural Networks using this small dataset and propose that such technique will work even better when there are millions of instances in the dataset.

In the dataset, some features have small values, e.g., in the range 0–10, and some features have large values as in the range of 100–1000. This can make the learning process slow. Therefore, data are normalized to enhance fast learning in the machine learning algorithms. After normalization, the values in the dataset will be in the range of 0..1. The formula of normalization is expressed in equation (1), where X_{max} is max value of all the samples for a given feature, X_{min} is the Min value of all samples by the feature, and Xi is is the current value of the feature. After normalization, the values in the dataset will be in the range of 0 to 1:

$$X_i = (X_i - X_{min}) / (X_{max} - X_{min}) \quad \text{Eq. 1}$$

3.3 Performance evaluation
The classification algorithms are ranked based on its testing accuracy and less computational complexity. Accuracy [2] of a classifier is measured in terms of how precisely the classifier places the input datasets under the correct category [2, 3]. This is denoted as the Misclassification rate which is computed as 1- Accuracy(C) where C denotes Classifier.

We have 740 samples in the dataset. We divide it into 503 training sample and 200 validating sample and 37 samples testing then we imported the dataset in Just Neural Network (JNN)[22] environment (as shown in Figure 1). We then trained and validated the JNN model (as seen in Figure 2). We found the most important attributes contributing to the JNN model as shown in Figure 3. The details of JNN model is shown in Figure 4. The Architecture of the JNN model we used consists of 5 layers: one Input layer, three hidden layers, and one output layer as shown in Figure 5. The hidden layers consist of (8 x 1 x 1) nodes. The controls of the parameters of the model are shown in Figure 6. The accuracy of the ANN model was 99%. Then we tested the model with the 37 samples kept aside and we found the accuracy to be 99% also.

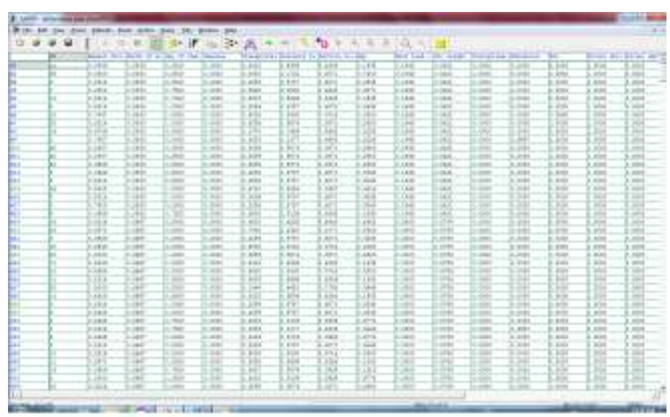


Figure 1: Imported the data set in JNN environment

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 9, September – 2020, Pages: 62-67

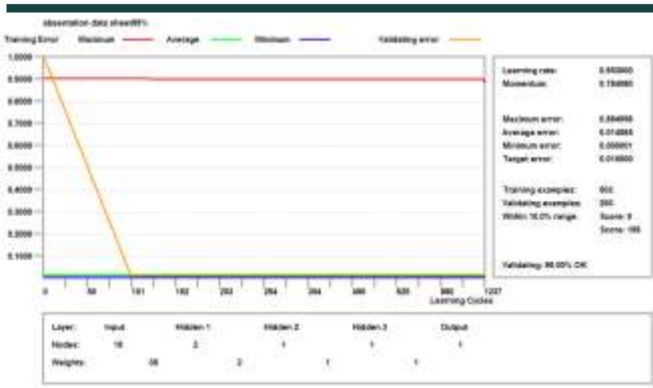
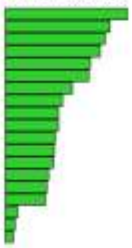


Figure 2: Training and validating of the ANN model in JNN environment



absentation data sheet99% 1237 cycles. Target error 0.0100 Average training error 0.014066
The first 19 of 19 inputs in descending order.

Column	Input Name	Importance	Relative Importance
3	Day of the week	0.9634	
1	Reason for absence	0.8233	
9	Work load Averageday	0.7825	
6	Distance from Residenc+	0.7491	
8	Age	0.6679	
15	Social smoker	0.6583	
5	Transportation expense	0.5315	
10	Hit target	0.4668	
7	Service time	0.4229	
16	Pet	0.4213	
4	Seasons	0.3950	
19	Body mass index	0.3930	
17	Weight	0.3825	
18	Height	0.3472	
2	Month of absence	0.3330	
14	Social drinker	0.3195	
11	Disciplinary failure	0.1025	
13	Son	0.0884	
12	Education	0.0726	

Figure 3: Most important attributes of the ANN model

Figure 4: Details of the ANN model

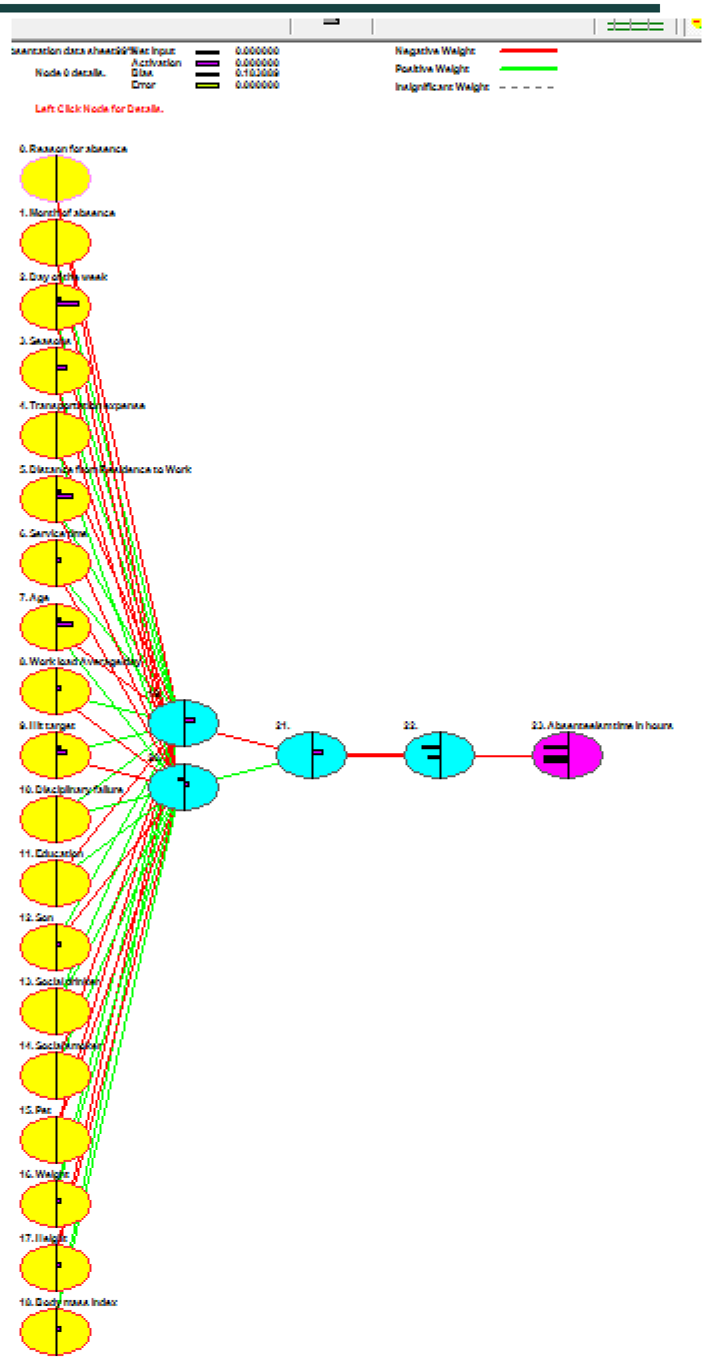


Figure 5: Architecture of the ANN model

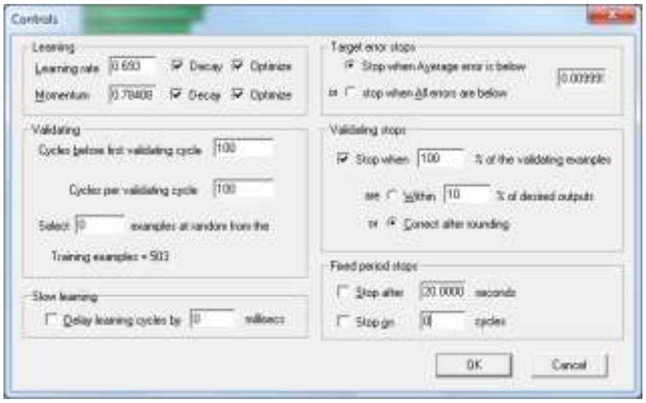


Figure 6: The controls of the parameters of the ANN model

4. Conclusion

Associations are very concerned about the behavior of employees towards punctuality at work, as associations can make progress only if employees are putting enough number of hours to work. But those employees who are habitually late and looking for excuses to steal time from work can be a real trouble for Associations. In order to avoid the critical issue of absenteeism where Associations have to confront those who are stealing time from work, a structured methodology using machine learning and just neural network model has been presented which can determine the behavior of such employees towards punctuality at work at the very early stage of their hiring. The results obtained have demonstrated higher accuracy (99%) by the proposed model and show a great potential to be scaled by using a big dataset for real-world problems. The dataset used in this research study belongs to a courier company in Brazil, and the dataset contains features which reflect human behaviors; such behaviors may correlate differently in other geographic locations. The proposed model can be extended to other global locations by adapting to local employee's behavioral features. Then; the cultural differences and demographic issues will not affect the model efficacy.

References:
1. Sturman, M. C. (1996). "Multiple Approaches to Absenteeism Analysis, Cornell University", School of Industrial and Labor Relations, Center for Advanced Human Resource Studies, London, UK.
2. Cucchiella et al. (2014). "Managi`ng absenteeism in the workplace: the case of an Italian multiutility company," Procedia-Social and Behavioral Sciences, vol. 150, pp. 1157–1166, 2014.
3. Roelen, e. al. (2009). "Trends in the incidence of sickness absence due to common mental disorders between 2001 and 2007 in the Netherlands," ?e European Journal of Public Health, vol. 19, no. 6, pp. 625–630, 2009.
4. Russo, et al. (2013). "How time and perceptions of social context shape employee absenteeism trajectories," Journal of Vocational Behavior, vol. 83, no. 2, pp. 209–217, 2013.
5. Gayathri, "Data mining of absentee data to increase productivity," International Journal of Engineering and Techniques, vol. 4, no. 3, pp. 478–180, 2018.
6. Ferreira, A. Martiniano, D. Napolitano, E. B. P. Farias, and R. J. Sassi, "Artificial neural network and their application in the prediction of absenteeism at work," International Journal of Recent Scientific Research, vol. 9, no. 1, pp. 2332–2334.
7. Trivedi, Explaining Absenteeism at Workplace Predicted by a Neural Network, Springer, Berlin, Germany, 2010.
8. Martiniano, R. P. Ferreira, R. J. Sassi, and C. Affonso, "Application of a neuro fuzzy network in prediction of absenteeism at work," in Proceedings of the 7th Iberian Conference on Information Systems and Technologies (CISTI 2012), vol. 1–4, Madrid, Spain, June 2012.
9. Shandizi, Prediction of Pilot's Absenteeism in an Airline Company, Universite De Montreal, Montreal, Canada, 2014.
10. Qomariyah and Y. G. Sucahyo, "Employees' attendance patterns prediction using classification algorithm case study: a private company in Indonesia," Int'l Journal of Computing, Communications & Instrumentation Engg, vol. 1, no. 1, pp. 2349–1477, 2014.
11. Punnoose and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," International Journal of Advanced Research in Artificial Intelligence, vol. 5, no. 9, 2016.
12. Ribes, K. Touahri, and B. Perthame, "Employee turnover prediction and retention policies design: a case study," CoRR, vol. 10, 2017.
13. Rohit and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," International Journal of Advanced Research in Artificial Intelligence, vol. 5, p. 10, 2016.
14. Esmaieeli Sikaroudi, R. Ghousi, and A. Sikaroudi, "A data mining approach to employee turnover prediction (case study: arak automotive parts manufacturing)," Journal of Industrial and Systems Engineering, vol. 8, no. 4, pp. 106–121, 2015.
15. Halland, C. Igel, and S. Alstrup, "High-school dropout prediction using machine learning: a Danish largescale study," in Proceedings of the ESANN 2015, M. Verleysen, Ed., pp. 319–324, i6doc.com, Bruges, Belgium, 2015.
16. Lakkaraju, E. Aguiar, C. Shan et al., "A machine learning framework to identify students at risk of adverse academic outcomes," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, ACM, New York, NY, USA, 2015.

17. Cortez and A. Silva, Using Data Mining to Predict Secondary School Student Performance, EUROSIS, London, UK, 2008.
18. Wahid, A. K. M. Z. Satter, A. Al Imran, and T. Bhuiyan, “Predicting absenteeism at work using tree-based learners,” in Proceedings of the 3rd International Conference on Machine Learning and Soft Computing, ICMLSC 2019, ACM, New York, NY, USA, 2019.
19. C. W. Kang, M. Imran, M. Omair, W. Ahmed, M. Ullah, and B. Sarkar, “Stochastic-petri net modeling and optimization for outdoor patients in building sustainable healthcare system considering staff absenteeism,” Mathematics, vol. 7, no. 6, p. 499, Jun 2019.
20. A. Tonge and C. Caragea, “Image privacy prediction using deep features,” in Proceedings of the ?irtieth AAAI Conference on Artificial Intelligence, AAAI’16, pp. 4266-4267, AAAI Press, New York, NY, USA, 2016.
21. UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)
22. JustNN Tool. http://www.justnn.com