Predicting performance analysis of garments women working status in Bangladesh using machine learning approaches

Conference Paper · February 2021

DOI: 10.1109/ICICT50816.2021.9358696

6 authors, including:

Maria Sultana Keya
Jahangirnagar University
10 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Minhaz Uddin Emon
Jahangirnagar University
10 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Md. Al Mahmud Imran
Daffodil International University
12 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Mayen Uddin Mojumdar
Daffodil International University
3 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Predicting performance analysis of garments women working status in Bangladesh using machine learning approaches

Maria Sultana Keya*, Minhaz Uddin Emon†, Himu Akter‡,
Md. Al Mahmud Imran§, Md. Kamrul Hassan¶, and Mayen Uddin Mojumdar‖
*†‡§¶‖Dept of Computer Science and Engineering
Daffodil International University, Dhaka-1207, Bangladesh
Email:*maria.sultana.keya@gmail.com, †minhazkhondokar21@gmail.com, ‡himu15-1212@diu.edu.bd,
§imran15-1243@diu.edu.bd, ¶hassankamrul801@gmail.com, ‖mayen.cse0285.c@diu.edu.bd

***Abstract*—In Bangladesh, the garment industry has played an important role in economically uplifting a diverse community of poor and marginalized people. There are now 4,825 garment factories that employ more than three million people. Completely 85% of these employees are female. But most of the female workers work to support their family and also contribute his family to lead a minimum life. In this paper, we try to find out relation between their health status, their family earning, their family member information, their working time or how many year they work in this sector and how many time they want to work. The dataset is collected from the Ashulia and Gazipur area garments of Bangladesh. This research work has observed that most of the female workers work at finishing, swing, helper, and cleaner sector. In this sector they cannot get huge salary that's why their income is limited and the range of their salaries is very low. It has also been found that, some women manage their whole family with their own income. Besides they are feeling bored with the same work. Nowadays machine learning and data mining tools play a vital role in finding the measurement of some important factors. This paper analyses the women working performance based on their previous activity and use some machine learning algorithms likely: Decision Tree Classifier(DTC), Logistic Regression(LR), Random Forest Classifier(RFC), and Stochastic Gradient Descent(SGD) we get the best result from Logistic Regression(LR) and it is 69%.**

***Keywords*—Bangladesh, Garments Women, Machine Learning, Cross Validation, Logistic Regression.**

I. Introduction:

Different mathematical and machine learning algorithms are now being applied in various areas, such as education, retail, health and medical problems, weather forecasting, study of socioeconomic activity, etc. Female Garments Worker data has also developed. From Bangladesh 's viewpoint, the number of female garment workers is much greater than the number of male garment workers. 4621(2018-19) [1] Garments Factories Source by BGMEA [2] is currently available in Bangladesh. Women also make up 90% of the 4 million staff in the Bangladeshi textile industry. A significant cohort of poor and vulnerable women have played an important role in economically uplifting the garment industry in Bangladesh. The chance to be working profitably has acted as a repellent against early marriage and in turn, decreased fertility. Income regulation also provides women with more power at home to make choices, a voice in the social sphere, and self-esteem. The social conditions of female garment workers in Bangladesh's Dhaka City are highlighted in this article. In rural areas, 90% of female employees are working. In this article, women in the textile factory/industry work an average of 11.12 hours / day but earn an average salary less than Tk. Per month, 7000. This study found that in a suppressor and unfavorable system, female clothing workers work. Their managers and their friends at work harass them. They are often sexually assaulted at the workplace where they work.

As an outcome, in the Female Garments Worker, the number of employees is much larger than the male Garments Employees. The textile industry in Bangladesh has made a revolution in the work of women in the manufacturing sector. The rapid growth in Bangladesh for the ready-made-garments (RMG) sector has contributed towards the availability of inexpensive and easily available women labour. As the country's biggest exporter, the hard work of millions of female workers has also reflected in the production of clothes. The garment industry's working climate, however, it is not safe, friendly and healthy for workers, particularly female employees. The human state of the workforce is related to some of the problems faced by women workers. This data uses information from the industry 's customers, buyers, and employees. A variety of patterns can be established by analyzing this data, which will help to make a decision. The industry currently employs 1,5 million workers in this sector, we are using different data mining and machine learning strategies on these data, approximately 80 percent of which are women, many of whom work in dangerous social environments. It has become a major source of employment for rural migrant women in a world where rural livelihoods have been steadily diminished and where migrant women have been largely removed from the situation. Women's workers provide cheap and easily exploited labor that helps the garment industry in Bangladesh to compete in the global market. Though studies have shown that women's jobs in Bangladesh's

export-oriented clothing industry has narrowed the gender gap in many respects, including participation in the labor force, social status, income control and decision-making, gender inequality remains prevalent in wage rates and social working conditions.

After the conclusion of this report, it will help to recognize the various methods to predict the involvement of employed girls in Bangladesh has steadily increased amid the difficulties and restrictions due to the lack of a friendly working climate. When women are involved, the problems of the working world are more significant. The key goals of this study are: to predict the female garment worker using different machine learning algorithms, considering difficulties and restrictions, female worker employment has steadily increased in Bangladesh due to the lack of congenial working climate of To classify which algorithm offers the best performance, different machine learning algorithms [3]. The rest of the paper as follow -

- Section I is briefly describe about garment worker known as introduction section.
- Section II is some previous work explanation name this related work.
- Section III research methodology describe briefly.
- Section IV experiment result analysis has given.
- Section V conclusion are describe here.

II. Related Work

Imran et. al. [4] performed a study on predicting the productivity of garment employees using deep neural networks (DNN). They evaluated the performance of the model by using MSE, MAE, MAPE metrics and hidden layers, neurons, optimization algorithms are being used in DNN. Interval between target and actual productivity. This model proved the unity from productivity data of garment employees.

Majumder et. al. [5] proposed a system that gender imbalances in export-oriented. They collect much information from the World Bank and BIDS. They used quantitative data and qualitative data. They eliminated the gender imbalances crop out from women's employment. Laws require to be purified to compromise equal sharing.

Lee et. al. [6] performed to achieve better quality assurance in the garment industry. They identified fuzzy association rule mining (FARM) for ascertaining quality assurance. They applied a genetic algorithm(GA) to hybridize FARM for optimization. Boolean association rule mining is being used for quality improvement. Insertion and deletion are two types of mutation. Rule generation module, rule optimization module and decision making module are being used. Restricted in sGAPMS. In future improve an intelligent system to assist QI operations. [7]. Khosla et. al. [8]identified to reduce social, political, and economic inhibition faced by women in the ready made garments industry. 23% of women are prime to the export processing zone aware of Bangladeshi's labor law.women are confess training from different NGOs. Latest date is not gainable and doesn't actual youthful statistical data. Needed to promote the cognition to protect women's rights.

Islam et. al. [9] performed to endeavour to analyse the contribution of RMG in economic development and challenges. They collected data from various journals, thesis papers , newspapers , survey reports and online news. They used primary data from BGMEA. Lacking many integration like poor inequality in training purposes.

Akhtar et. al. [10] proposed to find out the hazardous factors for improving occupational stress in industry. They gathered demographic and personal information. They collected primary data with occupational streets. Data is being collected by independent sample tests .To alleviate dangerous measures the association security.

Amin et. al. [11] identified to explore the indication of work as early socialization of young women. They used qualitative data and quantitative data for labor force participants to expand garments manufacturing to provide some setting of young Bangladeshi women. Illegal classes are defined in working space. Find out these irregularities and try to solve the problems.

Ahmed et. al. [12] found the social, political and economic affection and women earning Passion household gender dynamics within a framework of exit and voice. No surety in job safety. In vocation to the economic model. Begum et. al. [13] proposed that date was gathered personally. Quantitative and qualitative models are used so that to find true images. Date sets were not enough in their works and only 90 women's workers.

Devi et. al. [14]Proposed machine learning strategies in reviews based on numerous statistics and machine learning to solve the problem . For optimization sGenetic Algorithms (GA) and Particle Swarm Optimization (PSO) were being used.

Khare et. al. [15] performed balance in women labor in garments workers units. Used chi-square, T-test, percentage analysis to explore date. To pick various effects measures in the women workers . In future recognise the negative impact and try to solve the problem. Akhter et. al. [16] found to promote the total production of garment . The main factors to safety and health remittance. In industry's not sufficient courteous and healthful working environment. Try to stimulate ease and propose productiveness .

Ali et. al. [17] proposed establishment took space after involvement in garment. Quantitative methods are used. This paper carries to measure the exchanging establishment. Restricted structure to commission in health of government laws. This paper helps to inspire commission in garment worker health by shifting tax discounts. Sikdar et. al. [18] described the socio economic agreement of women garment workers. To explore percentages of frequency, frequency distribution are being used. This paper conceives some solutions and developing establishments . Mishra et. al. [19] performed to appreciate the alternative process.Fuzzy association rule, regression, Decision tree are used and k-means, k- nearest,cluster used. Making up the computable methods by traditional forecasting.

TABLE I
COMPARATIVE ANALYSIS OF EXISTING WORK AND OUR WORK

Existing Work	Our Work
In previous all work they focus about garment industry material, production effectiveness, ratio of male and female work	In our work we create an analysis of garment working women different categories and we find out accuracy of their work running and letting status.

III. RESEARCH METHODOLOGY

A. Data Description:

The dataset is used which has been collected from some reputed garments factory. It contains 512 workers records. This research has 13 attributes one attribute is response variable and the others are predictor variable. Some of them are,

AgeCat: This is the age range of the worker's. Here we partition age into 4 part, for 1 category age range is 17-25, for 2 category age range is 25-35, for 3 category age range is 35-45, for 4 category age range is 4-55. In our dataset lowest age is 17 and highest age is 55.

Marital Status: It represents whether the worker is married, unmarried or divorced.

Any Health Problem: It represents whether she has any health problem or not.

Salary Type: The quality of her salary.

Continue Work: It represents how long she wants to work in this company.

Status: It represents that if she left this company or not. Only the continue work attribute's data type is numerical and the others are categorical. Status is decision class and other is response class.

B. Algorithms Description:

Decision Tree(DT) Classifier: Tree-like structures are used in Decision Tree classification [20]. The root nodes means conditions and child nodes means the class label. Branches of the root nodes means effects of the conditions. The Entropy E(S) can be represented as figure 1

Where the probability of the jth class is fi.

$$E(S) = \sum_{j=1}^{c} - p_j \log_2 p_j$$

Fig. 1. DT equation

Logistic Regression(LR): Linear Regression analysis can't be applied if the relationship is nonlinear [21]. In that case the Logistic Regression can be used. The equation of Linear Regression is in figure 2

$$y = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \ldots\ldots + \alpha_n Z_n$$

Fig. 2. LR equation 1

Here, y is the response variable and Z1,Z2,Z3,.....,Zn are the predictor variables. $\alpha 0 = Intercept, \alpha 1, \alpha 2, ...\alpha n$ = coefficient. Applying sigmoid function on the equation, we get the logistic function in figure 3.

$$l = 1/[1+e^{-(\alpha_0+\alpha_1 Z_1+\alpha_2 Z_2+\ldots+\alpha_n Z_n)}]$$

Fig. 3. LR equation 2

Random Forest(RF) Classifier: It is a learning algorithm based on an ensemble tree. A selection of decision trees from the randomly chosen training set subset is the Random Forest Classifier [22]. To evaluate the final class it takes votes from the different decision trees. We use the mean squared error which is MSE while using the Random Forest Algorithm to fix regression problems.

Here n = number of data points, fi = value that the model returns and yi = the actual value of the data. Here N = number of data points, fi = value that the model returns and yi = the actual value of the data.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2$$

Fig. 4. Rf equation

Bagging Classifier: The bagging classifier is an ensemble meta-estimator that fits each base classifier on the random subsets of the original dataset and then aggregates their individual predictions to form a final prediction [23]. It is usually used as a way to reduce the variance of a black-box estimator. Predictions for unseen samples x' can be made after training by combining the predictions on x' from all the individual regression trees, So, predicting function is in figure 5. Where B is the number of trees which is free parameter.

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

Fig. 5. Bagging equation

SGD Classifier: Stochastic Gradient Descent (SGD) is very effective approach to fitting convex loss functions with linear classifiers and regression. It needs number of hyper parameters [24]. The algorithm reviews the training examples. For each example it need to updates the model parameters according to the update rule which is in figure 6. Here η is the learning rate, w is weight vector which lies in the x,y plane, is the loss function of i-th observation.

C. Data Visualization:

In this segment we figure out some bar chart which provide a graphical representation of our data set in

$$w \leftarrow w - \eta \left[\alpha \frac{\partial R(w)}{\partial w} + \frac{\partial L(w^T x_i + b, y_i)}{\partial w} \right]$$

Fig. 6. SGD equation

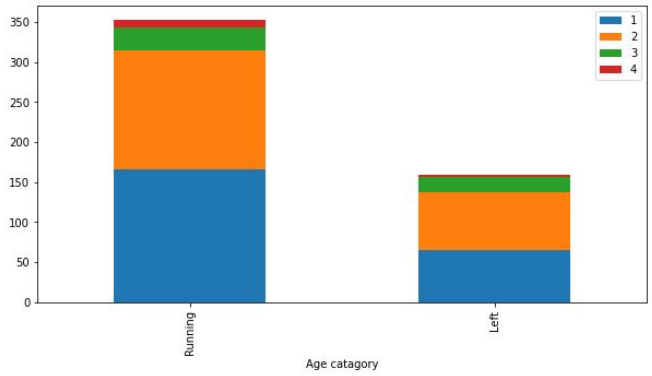

Fig. 7. Age category bar chart with respect to status attribute:

In figure 7 we categories age by 4, portion 1 is age range 17-24 years, 2 is 25-34 years, 3 is range 35-55 and above 55 year plus we categories as 4.

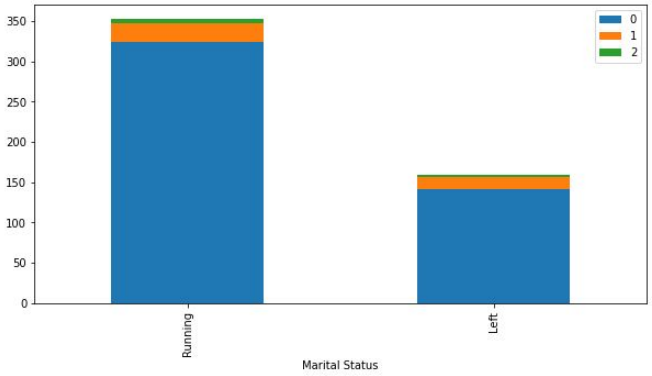

Fig. 8. Marital status bar chart with respect to status:

In figure 8 we categories Marital status by 3, portion 0 means Married, 1 means Unmarried and 2 means Divorced.

In figure 9 we categories Any health problem by 2 portion 0 means No and 1 means Yes.

In figure 10 we categories Family earning by 2 portion 0 means My income and 1 means Other

In figure 11 we categories Salary type by 4 portion 0 means Below Average, 1 means Average, 2 means Good and 3 means Excellent.

In figure 12, we found that women how many years they are willing to work. In figure right number of years given.

D. Data Processing:

In Figure 13 we draw a graphical representation of proposed model processing.

Input Data: We take 512 data of garments women for

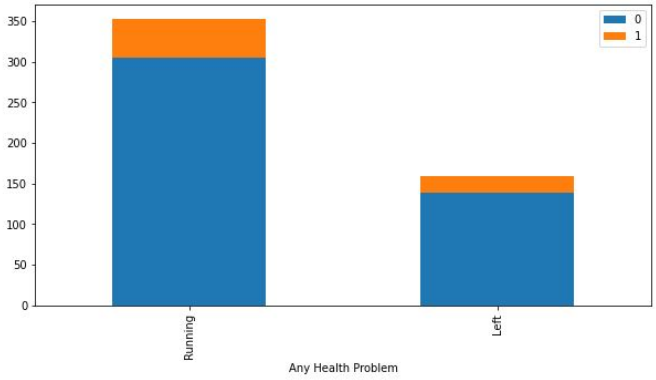

Fig. 9. Any health problem bar chart with respect to status:

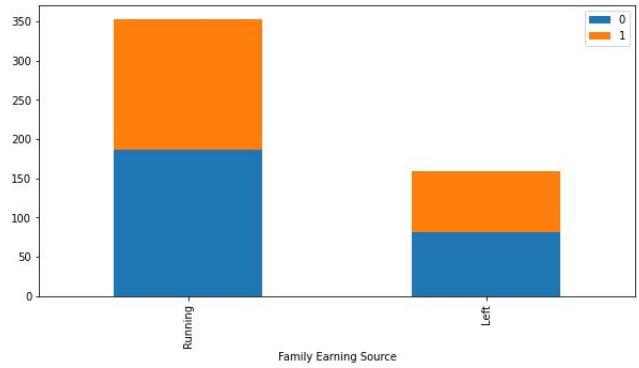

Fig. 10. Family earning bar chart with respect to status:

evaluation. This data collection form direct questions among garments women from different industry like as: Marma, Tusuka, Nittex industries limited, Islam group, Florance group, Fuljhuri industries imited, Sm sourcing, Kasham lamps limited and many more in Ashulia and Gazipur area in Bangladesh

Data Prepossessing: Data prepossessing step is categorized into two categories, namely: Data normalization and Encoding the categorical data into numeric data.

Data Visualization: In this segment we figure out some bar chart which provide a graphical representation of our data set in Age, Marital status, Any health problem, Family earning, Salary type, Continue work.

Ml Classifier: We use five machine learning classifier to train and test our data set. We use Decision Tree Classifier, Logistic Regression, Random Forest Classifier, Bagging Classifier, and SGD Classifier as our base algorithms to predict accuracy. After training all the classifiers, then we test our model.

Cross Validation: In this step we get the best accuracy from Logistic Regression, We specify how many folds we want to split in our data set. Here, we used 10-fold cross validation (k=10) to split the data into ten folds. We print out training indexes in each iteration to see clearly the K-Fold cross validation process.

Best Accuracy: We performed different machine learning classifiers to predict performance of women workers in the

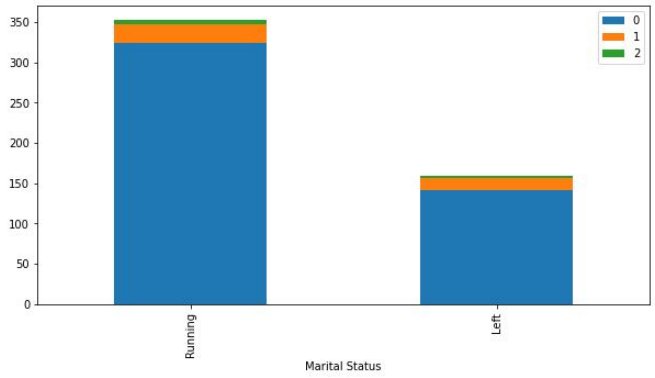

Fig. 11. Salary type bar chart with respect to status:

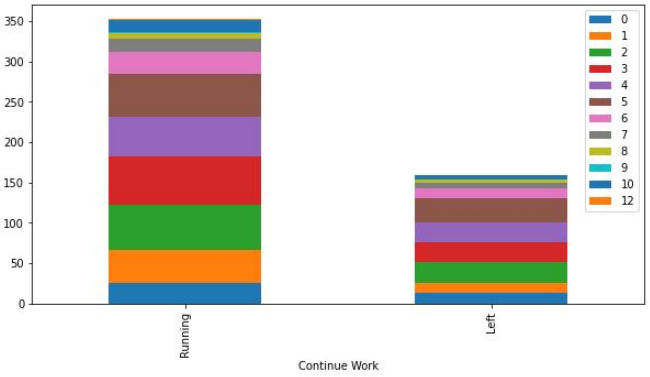

Fig. 12. Continue work bar chart respect to status:

garment industry.Like, from Decision Tree Classifier we get 63% accuracy, Logistic Regression we get 69%, Random Forest Classifier we get 62%, Bagging Classifier we get 61%, SGD Classifier we get 50%. We achieve 69% accuracy from Logistic Regression it is best accuracy from all other classification.

Best Model: After performing these model for this data set we get best accuracy from Logistic Regression. So, the best model is Logistic Regression among these model.

IV. Experiment Result Analysis:

In this paper, we performed different machine learning classifiers to predict performance of women workers in the garment industry. Here we get the best accuracy from Logistic Regression, we applied k-fold cross validation to evaluate our accuracy, Lowest accuracy we get from SDG Classifier. We achieve 69% accuracy from Logistic Regression it is best accuracy from all other classification. All the model that we use in this research to find out best accuracy, that's why we use multiple machine learning model to evaluate our accuracy. In this paper all this model like as: Decision Tree Classifier(DTC), Logistic Regression(LR), Random Forest Classifier(RFC), and Stochastic Gradient Descent(SGD) are base machine learning approaches that we use too evaluate our performance.In this paper we get 69% accuracy which is little bit low but our applying attribute has unique value many more that's the reason to get less accuracy. In this paper, we find out correlation metrics in figure 14 among feature variable to target variable.

TABLE II
Performance Analysis Of Five Classifier

Model Name	Accuracy	K fold(cross validation)
Decision Tree Classifier	63	0.63461538, 0.55769231, 0.60784314, 0.62745098, 0.7254902, 0.58823529,0.7254902, 0.49019608, 0.66666667, 0.66666667
Logistic Regression	69	0.65384615, 0.59615385, 0.78431373, 0.70588235, 0.78431373, 0.66666667, 0.74509804, 0.60784314, 0.66666667, 0.68627451
Random Forest Classifier	62	0.59615385, 0.57692308, 0.56862745, 0.64705882, 0.68627451, 0.60784314, 0.70588235, 0.50980392, 0.64705882, 0.66666667
Bagging Classifier	61	0.63461538, 0.53846154, 0.50980392, 0.58823529, 0.7254902, 0.56862745, 0.68627451, 0.54901961, 0.62745098, 0.62745098
SGD Classifier	50	0.65384615, 0.59615385, 0.19607843, 0.31372549, 0.66666667, 0.58823529, 0.35294118, 0.60784314, 0.66666667, 0.31372549

V. Conclusion:

In this paper, we want to create a relation with some attributes like a workers age, marital status, health problems and how many years they are willing to work. We generate a bar chart to graphical representation of their continued work with this sector for running and lifting status. We have been achieved that 69% accuracy result of their performance and We saw that most women working at garments to support their family and also improve their living condition. In future this study will work about garment products and women performance at product related issues and also use deep learning technique to achieve best result.

References

[1] Wikipedia contributors. (2020, October 13). Textile industry in Bangladesh. In Wikipedia, The Free Encyclopedia. Retrieved 19:12, October 29, 2020.

[2] Wikipedia contributors. (2020, April 20). Bangladesh Garment Manufacturers and Exporters Association. In Wikipedia, The Free Encyclopedia. Retrieved 19:16, October 29, 2020.

[3] Zulfiker, M. S., Kabir, N., Al Amin Biswas, P. C., Rahman, M. M. Predicting Students' Performance of the Private Universities of Bangladesh using Machine Learning Approaches.

[4] Al Imran, A., Amin, M. N., Rifat, M. R. I., Mehreen, S. (2019, April). Deep Neural Network Approach for Predicting the Productivity of Garment Employees. In 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT) (pp. 1402-1407). IEEE.

[5] Paul-Majumder, P., Begum, A. (2000). The gender imbalances in the export oriented garment industry in Bangladesh. Washington, DC: World Bank, Development Research Group/Poverty Reduction and Economic Management Network.

Fig. 13. Flow chart of proposed work section.

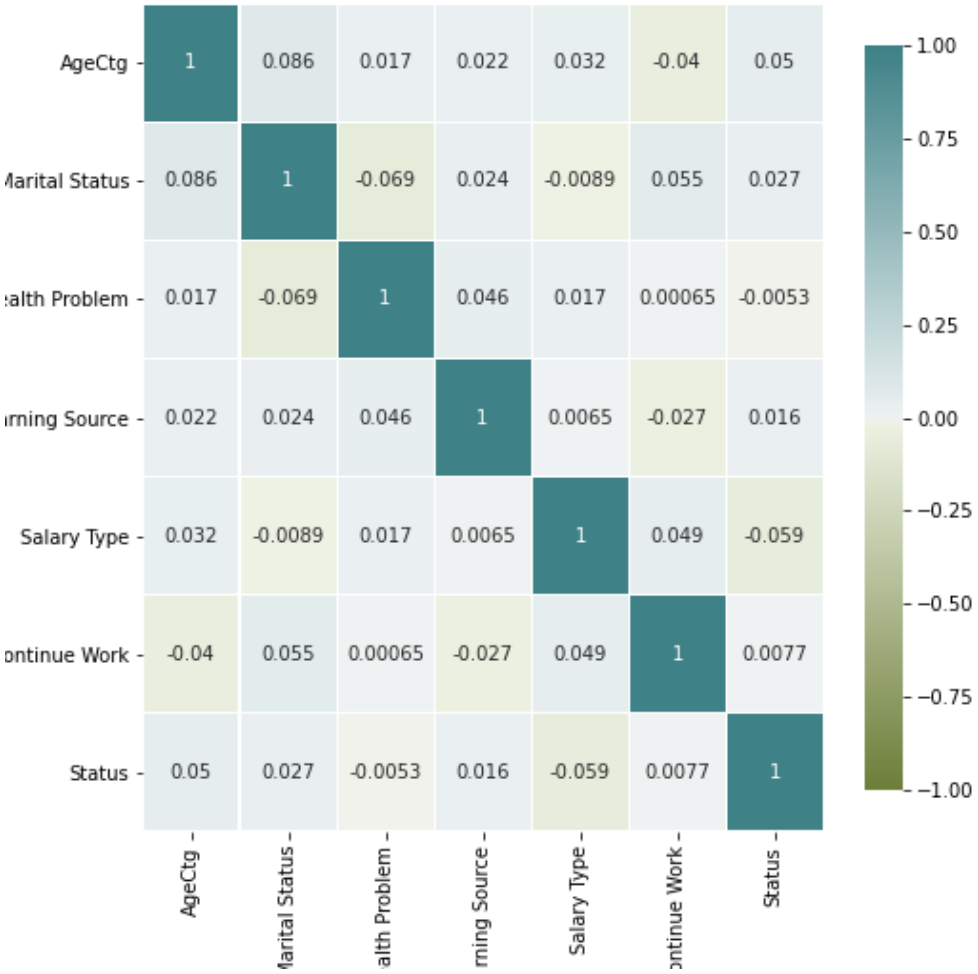

Fig. 14. Correlation matrices among the all attribute we applied.

[6] Lee, C. K. H., Choy, K. L., Ho, G. T., Lam, C. H. (2016). A slippery genetic algorithm-based process mining system for achieving better quality assurance in the garment industry. Expert systems with applications, 46, 236-248.

[7] Lee, C. K. H., Choy, K. L., Ho, G. T., Lam, C. H. (2016). A slippery genetic algorithm-based process mining system for achieving better quality assurance in the garment industry. Expert systems with applications, 46, 236-248.

[8] Khosla, N. (2009). The ready-made garments industry in Bangladesh: A means to reducing gender-based social exclusion of women?. Journal of International Women's Studies, 11(1), 289-303.

[9] Islam, M. S., Rakib, M. A., Adnan, A. (2016). Ready-made garments sector of Bangladesh: Its contribution and challenges towards development. Stud, 5(2).

[10] Akhtar, S., Shimul, A. M. (2012). Working hazards as indicator of occupational stress of industrial workers of Bangladesh. Asian Business Review, 1(1), 140-144.

[11] Amin, S., Diamond, I., Naved, R. T., Newby, M. (1998). Transition to adulthood of female garment-factory workers in Bangladesh. Studies in family planning, 185-200.

[12] Ahmed, F. E. (2004). The rise of the Bangladesh garment industry: Globalization, women workers, and voice. NWSA journal, 34-45.

[13] Begum, F., Ali, R. N., Hossain, M. A., Shahid, S. B. (2010). Harassment of women garment workers in Bangladesh. Journal of the Bangladesh

Agricultural University, 8(452-2016-35685)
[14] Devi, S. S., Radhika, Y. (2018). A survey on machine learning and statistical techniques in bankruptcy prediction. International Journal of Machine Learning and Computing, 8(2), 133-139.
[15] Khare, A. K., Kapoor, A. (2019). A Study on Work Life Balance among Women Labor in Garment Units of NCR, India. International Journal of Management, 10(1), 19-25.
[16] Akhter, S., Salahuddin, A. F. M., Iqbal, M., Malek, A. B. M. A., Jahan, N. (2010). Health and occupational safety for female workforce of garment industries in Bangladesh. Journal of Mechanical Engineering, 41(1), 65-70.
[17] Ali, R. N., Begum, M. M., Salehim, K. S., Farid, K. S. (2008). Livelihood pattern of rural women garment workers at Dhaka city. Journal of the Bangladesh Agricultural University, 6(452-2016-35506).
[18] Sikdar, M. M. H., Sarkar, M. S. K., Sadeka, S. (2014). Socio-economic conditions of the female garment workers in the capital city of Bangladesh. International Journal of Humanities and Social Science, 4(3), 173-179.
[19] Mishra, S., Mishra, D., Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: a review paper. Indian J. Sci. Technol, 9(38), 1-14.
[20] Wikipedia contributors. "Decision tree learning." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 13 Oct. 2020. Web. 16 Nov. 2020.
[21] Wikipedia contributors. "Logistic regression." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 5 Nov. 2020. Web. 16 Nov. 2020.
[22] Wikipedia contributors. "Random forest." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 25 Oct. 2020. Web. 16 Nov. 2020.
[23] Wikipedia contributors. "Bootstrap aggregating." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 21 Sep. 2020. Web. 16 Nov. 2020.
[24] Wikipedia contributors. "Stochastic gradient descent." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 2 Nov. 2020. Web. 16 Nov. 2020