International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ISSN: 2395-5325
Volume 3, Special Issue 3 (September '2017)
Proceeding of - International Conference on Recent Trends in Computational Life Science and Information Technology,
Conference held at Tiruppur Kumaran College for Women,Tirupur, Tamilnadu,India.

PREDICTING EMPLOYEE RETENTION USING DECISION TREES AND NEURAL NETWORKS

M.Kavitha,
Assistant Professor,
PG and Research Department Of Computer Science,
Tiruppur Kumaran College for Women,
Tirupur, Tamilnadu, India.

Jasmin Thomas,
Research Scholar,
PG and Research Department Of Computer Science,
Tiruppur Kumaran College for Women,
Tirupur, Tamilnadu, India.

Abstract: Rather being limited to pure academic research, Machine Learning Techniques and their practical applications is increasing their penetration into the main stream business processes. In business, correct decisions at the right time are quite important and have significant cost and risk implications. Decision Support Systems were developed for this where they assist and support key people to make business or organizational decisions. In this paper, we are discussing on the practical application of Machine Learning Techniques to predict the Employee Retention and optimize the hiring plan of an organization serving it as a decision support function for the Human Resource Management Systems. We have chosen Decision Forest and Artificial Neural Network as the algorithms and aim to discover the employee retention pattern in an organization.

Keywords: Human Resource Management Systems, Decision Forest, Artificial Neural Network, Azure Machine Learning Studio

I.INTRODUCTION

Retaining a positive and energetic workforce is very important for an organization's success. If the employee turnover is high, it increases the expenses and will have a negative effect on the company morale. Employee Retention is one of the major challenges faced by most of organization these days. So modern Human Resource Management System (HRMS) also act as a Decision Support System (DSS), integrating techniques from multiple areas including Machine Learning, Cognitive Science and even Behavioral patterns. This would help to hire better people and even predict possible employee attention. In this paper, we are discussing on using Machine Learning Techniques to empower an HRMS, so that it could result in better employee retention and optimize the organization resources. The models we are using here is a Two Class Decision Forest and a Two Class Artificial Neural Network with the sample data set of 15,000 employee data. The training was done using Azure Machine Learning Studio as it provides a way to do the experiment without any coding which could scale as required in terms of computing power and memory. Also, once the model is trained, we can easily publish the model into a web service which the HRMS software can consume it readily.

II. HUMAN RESOURCE MANAGEMENT SYSTEM

Human Resource Management System [1, 2] is described as an organization function which is responsible for recruiting, managing and directing people who are working with it. Main functions include: dealing with compensation, performance management, organization management, wellness, safety benefits, motivational training, grievance redressal and others. Even though there are a lot of functions for this department, we are specifically looking at the Employee Retention [6, 9, 10] aspect of the department. Employee Retention is the process by which the Employees are encouraged and motivated to remain in the organization for a maximum period of time, until the completion of the required task or Project. This is both beneficial for the organization and the employees. Researches show that there are series of four different expenses that may come up when a valued employee leaves. The first and most significant one is the financial cost, which will be particularly high for employees from crucial departments such as sales and key account managers. And added cost to that will be the cost for finding a new resource for that particular position. Replacing costs usually range from 30 to 200 percent of their annual salaries. The other two types of costs that are associated with turnover are loss in the knowledge when a long standing employee leaves and blow to the morale in the work environment.

III.PROPOSED FRAMEWORK

We will be using a Two Class Decision Forest [12] and a Two Class Artificial Neural Network [13] module to create two machine learning models. Even though there are several algorithms suited for the same problem, we are more aligned with the Decision Trees and Artificial Neural Network as Decision trees are fast, supervised ensemble model and Artificial Neural Network seems to be the answer for dynamic and complex problems. Main goal of the experiment will be the prediction of two values of Retention using two different models and comparing their results.

Decision Forests

Decision Forest [3, 4, 5] Algorithm is a great learning method for classification. It works by building multiple decision trees inline and ranking them on the most popular output class from the various features. Ranking is a form of aggregation where in which each tree which was built in a classification decision forest maps to a non-normalized frequency histogram of labels. The end result is derived from the aggregation process where we calculate the sum of

these histograms and normalizes the result to get the probabilities for each label. Trees with high prediction confidence will have a greater weight in the final decision. A sequence of simple tests is run for each class in each tree, by increasing the levels of a tree structure until a leaf node is met. Decision trees have many advantages compared to other classifiers:

- Symbolizing non-linear decision limitations
- Competent in computation and memory handling during training and prediction.
- Able to execute integrated feature assortment and categorization.
- In the presence of noisy features they are flexible
- Decision Forest classifier contains in itself a collection decision trees. Usually, ensemble models offer improved coverage and accuracy when compared to single decision trees.

Artificial Neural Networks

Artificial Neural Networks [7, 8] is a model which got inspired from the human brain where we try to model a network of interconnected neurons. Major function of each neuron is to propagate the input into the other neurons. This is usually done by a Transfer Function or Activation Functions. Every Neural Network will have many layers starting from two. The first and last layers are the input and output and other layers are called Hidden layer.

Activation Function or Transfer function plays a major role in the precision of the Neural Network. The mostly commonly used are Unipolar Sigmoid Function, Bipolar Sigmoid Function, Hyperbolic Tangent Function and Conic Section function. Similar to human brain the model tries to learn from the past experience and tries to use that to act on the future values. Here practically it turns down to find the training parameters called Weights. Neural Networks have several advantages when we compare to other classification algorithms

- The ease with which it can be handled.
- Able to estimate any function, despite of its linearity
- Great for complex/abstract problems like image recognition
- Mimics the brain & conceptually beautiful.

Implementation

To conduct an experiment, we have multiple options like coding it from scratch or using a tool. In the search for a tool, we ended up in Azure Machine Learning Studio [11].Exposure to Microsoft technologies was one of the key reason behind this decision. But technically it was a better tool in terms of power and flexibility. Desktop tools like WEKA are limited with the system memory or the processing power of the computer in which it is running. Microsoft Azure Machine Learning Studio a cloud based tool which had the benefit of scaling in terms of RAM and Computing power as required. Also it's built in cloud services, will help us to publish the trained model easily has a web service which could be integrated to the existing HRMS software.

IV. EXPERIMENTAL SETUP

It is easy to create a machine learning model in Microsoft Azure Machine Learning Studio. It is a drag and drop, collaborative tool which we can learn fast and use with less effort in a short period of time. It even has a free version to build, test and deploy machine learning experiments. To build the experiment, simply drag and drop the required components from the toolbox and connect the relevant steps. For the models we were trying, it was a fairly simple experiment that takes HR data in uploaded csv file and trains two different models. Of the two one was the Two-Class Decision Forest Classifier and the other one was Two-Class Artificial Neural Network. Final model looked like Fig.1

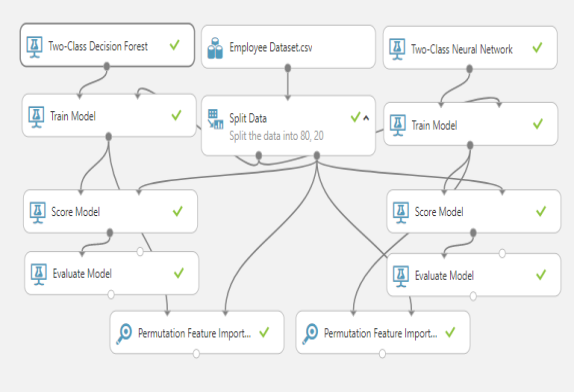

Figure1: Screenshot of the Experiment in Azure ML Studio

Training Data Set

Finding a right dataset is the first half of the problem. Recently kaggle published an open dataset for Human Resource Analytics. This had the details of about 15,000 employees. For every employee we have several attributes like satisfaction level, last evaluation, department, salary, number of projects etc. Dataset was divided into two parts (train = 80%, test = 20%) and models were trained on 80% dataset (training dataset) and evaluated on 20% (test dataset). Fig.2 and Fig.3 shows the accuracy metrics for each model

True Positive	False Negative	Accuracy	Precision
680	30	0.987	0.986
False Positive	**True Negative**	**Recall**	**F1 Score**
10	2280	0.958	0.971
Positive Label	**Negative Label**		
1	0		

Figure 2: Accuracy metrics for the Decision Forest model

True Positive	False Negative	Accuracy	Precision
634	76	0.954	0.911
False Positive	**True Negative**	**Recall**	**F1 Score**
62	2228	0.893	0.902
Positive Label	**Negative Label**		
1	0		

Figure 3: Accuracy metrics for the Neural Network model

The accuracy levels of both the models were more than 95% which means the models were performing well. Decision Forest had a higher accuracy compared to Neural Network for this particular experiment.

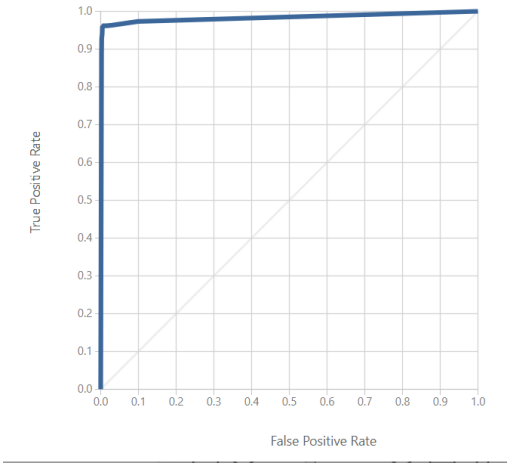

Figure 4: ROC curve for the Decision Forrest

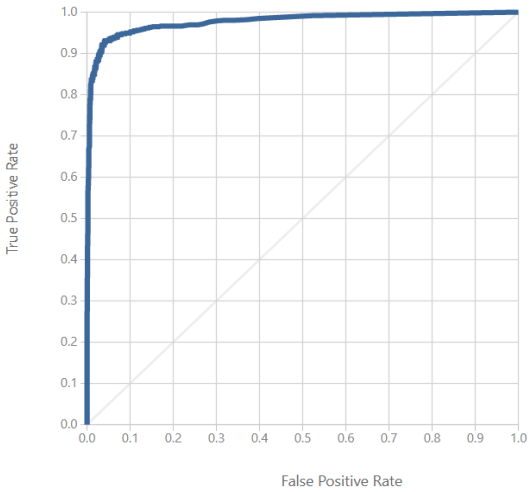

Figure 5: ROC curve for the Neural Network

V. CONCLUSION

Futuristic HRMS should also act as a Decision Support System where it would help the HR Managers to have a better understanding of the big picture on the grounds of Employee satisfaction levels. Access to this information will help the Managers to provide the best in class environment for the employees and to iterate things to improve it further. This will also help in reducing the cost associated with these problems and improve the morale value of the company culture. The current work shows that, with an ample amount of data points and using Machine Learning techniques we could build a decision support system which would help HRMS to a great extent.

VI. FUTURE WORK

The future scope could be like to extending the experiment into more algorithms like Support Vector Machine, Regression techniques etc and expanding the binary classification to multiclass classification.

VII.REFERENCES

[1]. Keegan A and Francis H, "Practitioner talk: the changing textscape of HRM and emergence of HR business partnership", The International Journal of Human Resource Management, 21(6), pp 873-98, 2010.
[2]. Schneider B, "The people make the place", Personnel Psychology, 40 (3), pp 437-53, 1987.
[3]. G.R. Dattatreya, L.N. Kanal, "Decision Trees in Pattern Recognition", Progress in Pattern Recognition 2, 1985.
[4]. T.K. Ho, "Random Decision Forests", Proc. Third Int'l Conf. Document Analysis and Recognition, pp. 278-282, 1995-Aug.-14-18.
[5]. S.W. Kwok, C. Carter, "Multiple Decision Trees", Uncertainty in Artificial Intelligence, vol. 4, pp. 327-335, 1990.
[6]. C.R. Coombs, "Improving retention strategies for IT professionals working in the public sector", Information and Management, vol. 46, pp. 233-240, 2009.
[7]. Pai G.A. Vijyalakshmi, S. Rajasekaran, "Neural networks fuzzy logic and genetic algorithms Synthesis and applications", Reading, Prentice-Hall of India, 2004.
[8]. D.E. Goldberg, "Genetic algorithms in search optimization and machine learning", Pearson Education, 2001.
[9]. BL Mak, H Sockel, "Info. & Mgmt", A confirmatory factor analysis of IS employee motivation and retention, 2001.
[10]. Augustine Agho, Charles W Mueller, Price, L James, "Determinants for employee job satisfaction: An empirical test of a causal model", Human Relations, vol. 46, no. 8, pp. 1007-1018, 1993.
[11]. Microsoft Azure Machine Learning Studio, [online] Available: https://studio.azureml.net/.
[12]. Two class Decision Forest, [online] Available:
[13]. https://msdn.microsoft.com/en-us/library/azure/dn906008.aspx.
[14]. Two-Class Neural Network, [online] Available: https://msdn.microsoft.com/en-us/library/azure/dn905947.aspx