Is Deep Auto encoder a Better Forecasting Tool? - Case of Employee Attrition

Kalyan Sankar Sengupta
Professor of Business Analytics
IFIM
B-School,Bangalore
Kalyan.sengupta@ifim.edu.in

K.T. Mithuna
PGDM Second Year
IFIM Business School, Bangalore
mithunakeerthi@gmail.com

T. Aashika
PGDM Second Year
IFIM Business School, Bangalore,
aashika.t@ifimbschool.com

***Abstract*- Human Resource (HR) manager is responsible for extracting the useful information or pattern from the past, present and future employees of an organization in order to retain the best employees. The behavior and intelligence of an employee can be analyzed in many ways using statistical models, mathematical models and business intelligence techniques. In fact, all the organizations or industries are having good database about the details of employees. For any organization, the role of employee is inevitable and the most of the companies are using innovative ideas to retain the recruited people. Nowadays, with the existence of Data Science and prediction techniques, this task can be automatically done, which allows the managers of the companies to obtain the information they require from the employees in a much faster and efficient way than it was obtained in the past when the task was done manually by the human resources department. These results in a significant decrease of the costs associated with employee attrition, turnover or churn, maximizing the revenue of the company. Many researchers have explored the data mining techniques such as decision tree algorithm, Back Propagation Neural Network, and Random Forest method. In this paper, a novel method based on machine learning has been proposed to predict the employee attrition using deep autoencoder. Deep Autoencoder consists of two halves, one representing the encoding half and the other representing the decoding half. The model has been validated is compared with the results of Naïve Bayes, Support Vector Machine (SVM), Back Propagation Neural Network (BPN) and Random Forest.**

Keywords - Attrition, Deep Autoencoder, BPN, SVM, Random Forest, Human Resource

1. INTRODUCTION

Over the past few decades researches evident that the productivity of an organization is highly dependent on their employees. In a highly competitive global market, employee attrition poses great threat and challenges for organizations. Generally, it is the responsibility of Human resource manager to pay considerable attention on employee attrition and its impact. In the digital era, Predictive analysis allows the analyst to operate on historical and current information as well as predicting the likely future environment. This predictive insight promotes much better decision making and improved results. Use of predictive analytics is wide, it enables companies to improve almost every aspect of their business (Frye, Boomhower, Smith, Vitovsky, & Fabricant, 2018). One of the uses of Predictive Analytics for Human Resources is predicting employee attrition. Employee attrition has number of negative impacts including loss of enterprise knowledge, costs associated with leaving and replacement. To accurately determine who is leaving and what is the underlying reason are key issues for HR workforce planning. As Human Resources (possesses a massive amount of employee data, demand for analyses if high. However, HR Information Systems (HRIS) are often underfunded compared to information systems of other domains of enterprise, which are directly connected with main business (Boxall, Purcell, & Wright, 2007). This leads to the fact that HR data contains lot of noise and errors. Therefore, building accurate analytical model is challenging for HR.Analytical model of employee churn is presented in figure 1. First historical data should be collected of employee churn behaviour, and then it will be split into training and testing datasets. The dataset model will be built on training and validated on testing. Then current employee data will be analysed using the constructed model to make predictions. Predictive modeling for employee turnover faces few challenges. Such analytical model should be accurate and easily understandable for HR managers. They should be able to fully grasp drivers of employee dissatisfaction. Model for employee turnover should be operationally efficient, employees leave organizations on regular basis, therefore model should be updated at least monthly or quarterly. Outcome of the model, predictions should be used for employee retention campaigns selectively for key employees or for employees with high potential. Employee who does not perform well can be left out of targeted group. Depending on the reasons why an employee leaves an organisation, labor turnover can be categorized into two categories viz., voluntary and involuntary (Labov, Ash, & Boberg, 2008). In the voluntary turnover, the employee leaves the organisation of his own will. In involuntary turnover, other causes, which are not under the control of the employee takes place, which is initiated by the organization itself. This study emphasis on predicting voluntary churn employees. Moreover, it is proposed to predict employee attrition dataset provided by IBM. This study is useful for both industry and research perspective.

ISSN: 2349-6363

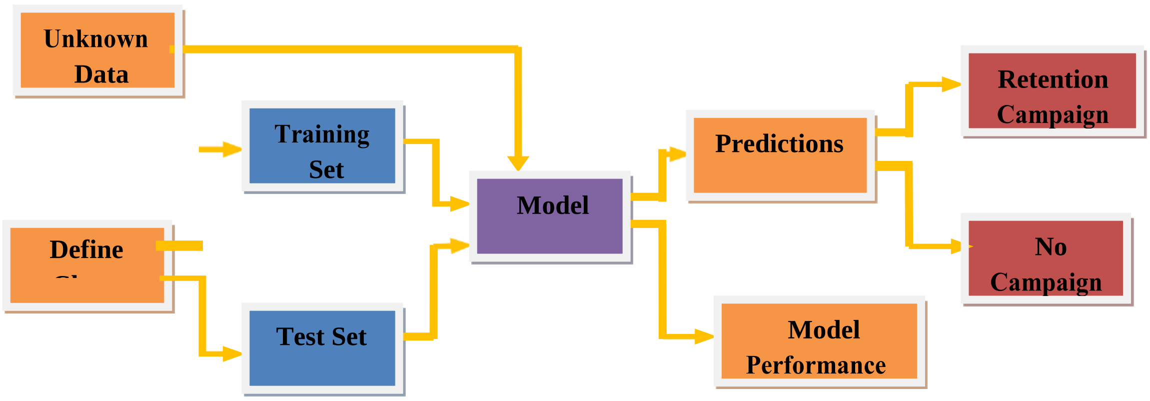

Figure 1. Analytical model of employee churn

In general, predictive analytics and machine learning go hand-in-hand to solve more complex real-world problems. Furthermore, Machine learning algorithms are often showcased in customer attrition prediction. Nowadays, a new machine learning model called deep learning is widely adopted for predictive analytics. They are more effectual in learning data representation. A salient feature of deep learning is that it allows a machine to be fed with raw data and to automatically discover the unique patterns requisite for further analysis. As a result, with the help of deep neural network, HR Managers can take numerous decisions on investment on employees to get the excellent outcomes that benefits the stakeholders and customers.

1.1 Motivation

The reason why organizations are putting so much effort into preventing employee turnover, i.e. employees leaving the company and having to be replaced, is that it has an adverse impact for the company, including a huge waste of time and money.

Employee attrition leads to losses in all these areas:

Overworked remaining employees: The rest of the employees have to take over the job that the person who has left was in charge of, which translates into unsatisfied employees, who are more likely to leave the company. Lost experience and knowledge: All the experience and knowledge that the former employee had acquired through the years are no longer available for the company. Recruiting costs: The selection process to replace the employee costs the Human Resources department both money and time. Training costs: In addition to the training courses the new employee might need, further costs must be taken into account. This worker will take some time to be fully able to do his/her job and he/she will need help from other co-workers, which leads to a significant drop in productivity.

1.2 Research Objectives

Research objective aims at predicting the employee attrition with deep learning models. Different predictive models are used to understand the attrition phenomenon. Once the attrition is found out, study on the factors deciding the valuable employees and after finalizing those factors, build the decision model for valuable employees. And then find the retention factors and display the most effective retention factors for each employee to improve the employee retention.

1.3 Major Contribution

It is proposed to develop a deep autoencoder model with multiple hidden layers which is trained layer by layer to predict the employee attrition in order to gain competitive advantage. Section 2 elaborately discusses the related work of employee attrition. In section 3, the proposed employee attrition prediction using deep autoencoder is presented. The experimental results are discussed in section 4. Finally, this paper concludes with some future perspectives in section 5.

2. LITERATURE REVIEW

Employee attrition prediction is one of the biggest challenging problems in human resource management literature, yet continues to elude any concrete conclusions. Many researchers argue that high attrition rates might have negative effects on the profitability of organizations if not managed properly (Saradhi & Palshikar, 2011). The actual costs have been estimated and published in a study by the Center of American Progress, which reveals that, in average, replacing a lost employee costs businesses one-fifth of the employee’s yearly salary (Boushey & Glynn, 2012). Demographical factors like age and location are strong predictors of employee turnover because the younger employees from age 18-25 are more likely to turnover than older employees. Since younger employees leave in early stages so in-position and in-service also have some effect on turnover. These results are consistent with a study on turnover rates conducted by Hill and Associates which found that young undergraduates, graduates and postgraduates in the outsourcing business had changed their jobs at least once in the past three years (Banerjee, 2008). The location in our analysis has an impact because most of the employees who leave the company are from UK.Globally competitive organizations will depend on the uniqueness of their human resources and the systems for managing human resources effectively to gain competitive advantage (Pfeffer, 1994) (Bartlett & Ghoshal, 1997) (Barney & Wright, 1998).(Booth & Hamer, 2007) Booth et al. found that labor turnover is related to a variety of environmental factors and organizational factors such as company culture and values, supervisory style, fair pay, corporate value, giving support to each other, trust and respect between employees, manageable workload, development and career building satisfaction and degree of job satisfaction. Another research done on the Indian IT firms reveals the factors responsible for the attrition are Below expectation salary, Low incentives, Relationship with superior, Relationship with subordinates, Job knowledge, Skills utilization, Skills recognition, Acknowledgement of work by superior, Lack of appreciation, Unsatisfied work culture (Archita Banerjee, 2017).One of the factors that help retention, is providing healthy work environment. Also, if reason of attrition is found out, retention can be improved through planning and involvement in discussion with employee (Frye, Boomhower, Smith, Vitovsky, & Fabricant, 2018). Huang et al. Proposes some new features to customer churn prediction and implement seven prediction techniques including Logistic Regression, Linear Classification, Naive Bayes, Decision Tree, Multilayer Perceptron Neural Networks, Support Vector Machines, and the evolutionary data mining algorithms (Huang, Kechadi, & Buckley, 2012).

3. EMPLOYEE ATTRITION PREDICTION USING DEEP AUTOENCODER

Recently, in the domain of predictive analytics, deep learning models are the promising avenues of research. The deep learning model automatically extracts complex representations from the raw data. More specifically, the deep model has the generalization ability to solve the real world problems. Empirical studies have explored that data representations attained from deep models often yield better performance (Wang, 2016). (Yosinski, Clune, Bengio, & Lipson, 2014) (Bengio, Courville, & Vincent, 2013). Before 2006, the neural network with more than one hidden layers was considered to be hard to train efficiently and gained popularity among the researchers when it was shown that training Stacked AutoEncoders (SAEs) layer-by-layer in an unsupervised manner, followed by supervised fine-tuning of the network in pattern recognition problems (Golovko, Kroshchanka, & Treadwell, 2016) (Singh & Om, 2017).

3.1 Stacked Autoencoder

Generally, at the core of all deep learning models is the domain-independent idea of utilizing hierarchical layers of learned abstraction to effectively accomplish high-level tasks. Among all the deep models, the deep autoencoder model has been applied to a wide variety of predictive problems with great success.

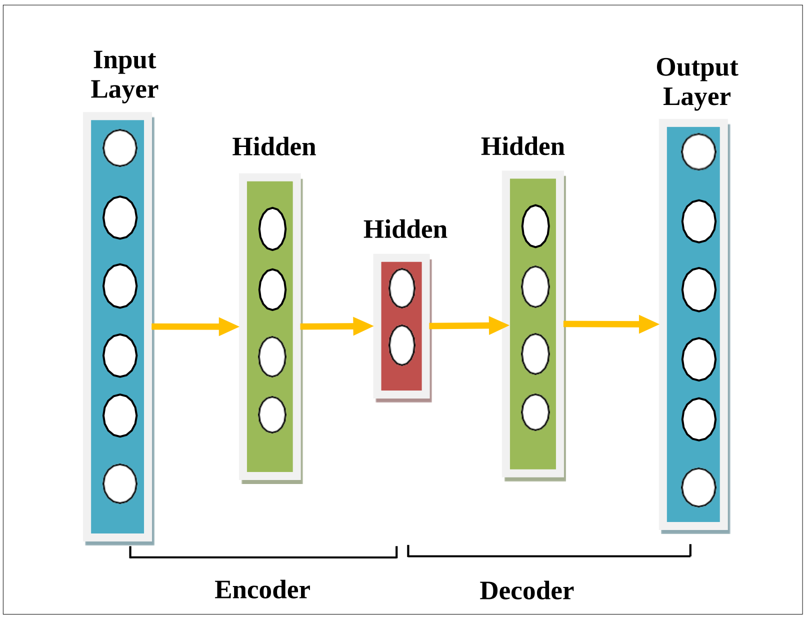

Figure 2. An autoencoder with 3 fully-connected hidden layers

More specifically, they are trained to encode the input x into some representation r(x) (Swersky, Ranzato, Buchman, Marlin, & de Freitas, 2011) (Vincent, A connection between score matching and denoising autoencoders, 2011). It consists of two parts, the encoder, and the decoder as shown in figure 2. The encoder part takes the raw data $I(x) \in R^d$ and maps it into a hidden representation $H(x) \in R^{'}$ through a deterministic mapping represented in equation (1).

$$H(x) = \sigma(W * I(x) + b \quad (1)$$

where σ is the transformation function used. Here, W represents the weight matrix and b is the bias value. The latent representation $H(x)$, or code is then mapped back into a reconstruction R of the same shape as $I(x)$. The mapping happens through a similar transformation as given in equation (2).

$$R(x) = \sigma(W^t * H(x) + b^l) \tag{2}$$

Here, R should be seen as a prediction of $I(x)$, given the encode $H(x)$. Autoencoders are trained to minimize reconstruction error between the raw input and reconstructed one (Vincent, Larochelle, Bengio, & Manzagol, 2008) (Rifai, Vincent, Muller, Glorot, & Bengio, 2011). The reconstruction error can be measured in many ways, depending on the appropriate distributional assumptions on the input given the code. The conventional squared error is given in equation (3).

$$L(I(x_t), R(x) = \|I(x) - R(x)\|^2 \tag{3}$$

If the input $I(x)$ is interpreted as either binary or vector of bit probabilities, then cross-entropy of the reconstruction can be used and is given in equation (4).

$$L(I(x), R(x) = -\sum_{k=1}^{d}[I(x)_k \log R(x)_k + (1 - I(x)_k)\log(1 - R(x)r_k \tag{4}$$

3.2. Attrition Prediction with deep autoencoder

The stacked autoencoder is a neural network consisting of multiple layers of sparse autoencoders, where the output of each layer is connected to the inputs of the successive layer. The schematic architecture of a deep autoencoder model with attrition prediction attributes to find the intrinsic relationship between predictor variables, then use the output (weight, encoded features) of the autoencoder model to initialize the classification model is presented in figure 3. In this study, softmax classifier is adopted for predicting the attrition.

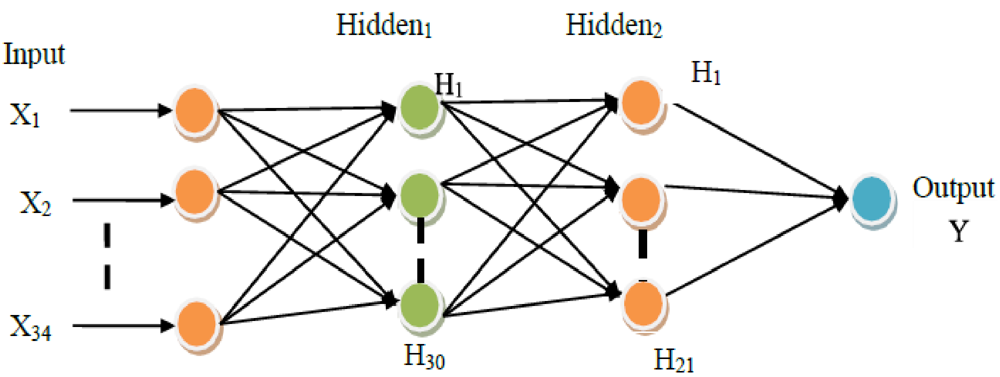

Figure 3. A schematic architecture of deep autoencoder for attrition prediction

Algorithm 1: Employee Attrition Prediction with Deep Autoencoder
Input: Attrition Dataset (n)
Output: Attrition: Y/N
Step 1: Encode the input to a hidden representation with an autoencoder through a deterministic mapping,

$$Encode\_1 = \sigma(W * n + b)$$

Step 2: The encoded information from step 2 is then given as an input to the second autoencoder for further compression through a deterministic mapping,

$$Encode\_2 = \sigma(W * Encode\_1 + b)$$

Step 3: the latent representation or code represents the optimal feature to the classifier
Step 4: Train a softmax layer for classification using the features from the second autoencoder
Step 5: The loss function (cross entropy) for the softmax layer is,

$$E = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} t_{ij}\, ln y_{ij}(1 - y_{ij})\ln(1 - y_{ij})$$

Here, n is the number of training samples, k is the number of classes, is the jth entry of target matrix, T, and is the ith output from the autoencoder.
Step 7: Evaluate the performance using precision, recall, f-measure, accuracy and error rate.

4. EXPERIMENTAL RESULTS

4.1. Dataset

The deep autoencoder model proposed for employee attrition prediction has been implemented and discussed. The experiments have been conducted by MATLAB on IBM HR employee attrition dataset created by IBM data scientists (IBM Attrition dataset). It contains 1470 employee records with 34 predictor variables and one response variable. The dataset also contains the various levels of education, job involvement, job satisfaction , performance rating which is categorized as segments-.The segments are numbered as 1 being 'low' and 4 being 'very high'.

4.2 Performance Measures

The performance measures such as precision, recall, f-measure, accuracy and error rate which are derived from the confusion matrix are used to test the constructed classifier. The performance of the system is examined by demonstrating correct and incorrect patterns as given in table 1.

Table 1. Confusion Matrix

Actual	**Predicted**	
	Positive	**Negative**
Positive	TP	FN
Negative	FP	TN

Table 2. Evaluation Measures

Metric	**Formula**
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F-Measure	$\frac{(2 * Precision * Recall)}{Precision + Recall)}$
Accuracy	$\frac{TP+TN}{TP+TN+FN+FP}$
Error Rate	$\frac{FP+FN}{TP+TN+FN+FP}$

Where TP represents True Positive, TN represents the True Negative, FP represents the False Positive and FN represents the False Negative (Sokolova & Lapalme, 2009).The quantitative metrics used are given in table 2. In general, the higher value of precision, recall, f-measure, and accuracy shows better performance of the system. Similarly, the lower value of error rate indicates the efficacy of the classifier.

4.3 Experimental Results

The efficacy of the proposed deep autoencoder model has been compared with the standard benchmark classifiers such as Naïve Bayes, SVM, BPN, and Random Forest.

4.3.1 Naïve Bayes

The Naïve Bayes classifier works on the principle of Bayes theorem with the naive assumption of independence between every pair of features in the pattern recognition problem (Leung, 2007). In particular, the Bayes theorem states the following relationship for a class y with a dependent feature vector $x_1,\ldots\ldots,x_n$:

$$P(Y|X_1, \ldots \ldots . X_n) = \frac{P(Y)P(X_1,\ldots\ldots X_n|Y)}{P(X_1,\ldots\ldots X_n)} \tag{5}$$

With the Naïve independent assumption eq. (3.14) is simplified

$$P(Y|X_1, \dots X_n) = \frac{P(Y)\pi_{i=1}^{n} P(X_i|Y)}{P(X_1, \dots X_n)} \tag{6}$$

Finally, the classification rule is defined as,

$$\hat{Y} = argmax_y P(y)\pi_{i=1}^{n} P(X_i|Y) \tag{7}$$

The Naïve Bayes classifier as one of the fast machine learning algorithm which is used in many pattern recognition problems.

4.3.2 Support Vector Machine

Generally, the Support Vector Machine defines the decision boundary based on the concept of decision planes which separates the objects having different class values (Phillips, 1999) (Yao, Marcialis, Pontil, Frasconi, & Roli, 2003). More specifically, the algorithm tries to find the optimum hyperplane which separates the objects with less classification error. The margin is given by,

$$\rho(w, b) = \frac{2}{\|w\|} \tag{8}$$

Here, w is the norm vector to the hyperplane and b is a constant. Hence the hyperplane that optimally separates the data is the one that minimizes,

$$\emptyset(w) = \frac{1}{2}\|w\|^2 \tag{9}$$

4.3.3 Back Propagation Neural Network

A neural network is a powerful tool to capture and represents complex relationships between input/output (Haykin, 2005). Furthermore, the influence of neural networks has been demonstrated in several pattern recognition applications in the past including biometric recognition, speech synthesis and analysis, diagnostic problems, medicine, business and finance, robotic control, signal processing, image processing and many more. The Back Propagation Neural network optimizes the net for correct responses to the training input data set.The number of hidden neurons is greater than or equal to the number of input neurons. Initial weights are assigned randomly. The output from each hidden neuron is calculated using the sigmoid function as

$$S_1 = \frac{1}{1+e^{x}} \quad \text{where } \lambda = 1 \text{ and } \quad \textstyle\sum_i W_{ih} k_i \tag{10}$$

where $_h$ is the weight assigned to the input and hidden layer and k is the input value of the network. The output is calculated using the sigmoid transfer function.

$$S_1 = \frac{1}{1+e^{x}} \quad \text{where } \lambda = 1 \text{ and } \quad \textstyle\sum_i W_{ho} s_i \tag{11}$$

where who is the weight assigned between hidden and output layer and S_i is the output value from hidden neurons. S_2 is subtracted from the desired output. Using this error value (E = Desired – Actual), the updating of weight is performed as:

$$\delta = ES_2(1 - S_2) \tag{12}$$

The weights assigned between input and hidden layer, the hidden and output layer are updated as:

$$W_{ho} = W_{ho} + (\text{n}.\delta.S_1) \tag{13}$$

$$W_{ho} = W_{ho} + (\text{n}.\delta.S_1) \tag{14}$$

where, n is the learning rate, δ is the weight and k is the input value to the network. Again the output is calculated for the hidden and output neurons. Then the error value (E) is checked, and the weights are updated. The above procedure is repeated till the target output is equal to the desired output.

4.3.4 Random Forest

In general, Random forest is an ensemble learning method that constructs multiple decision trees. CART can be used to construct the decision tree for prediction. In particular, CART recursively partitions on a nominal target category to reach a tree structure. As the decision tree grows, a feature must be identified to split on it. So, all features are compared to each other to select the best feature. This comparison can be done by the entropy or Gini index that measures purity of feature separation. The CART stopping rule occurs when the target feature in the last separations are insignificant. Each decision tree returns a class and then bagging combines them to reach a unique decision (Breiman, 2001).

Table 3. Comparative analysis of proposed deep autoencoder with existing benchmark classifiers

Classification Algorithm	Precision	Recall	F-Measure	Accuracy	Error Rate
Naive Bayes	70.10	71.10	70.40	71.14	28.86
SVM	85.30	85.50	85.00	85.47	14.53
BPN	87.70	89.90	87.60	87.86	12.14
Random Forest	89.00	89.00	86.00	89.00	11.00
Deep Autoencoder	**93.10**	**91.15**	**92.11**	**93.10**	**06.90**

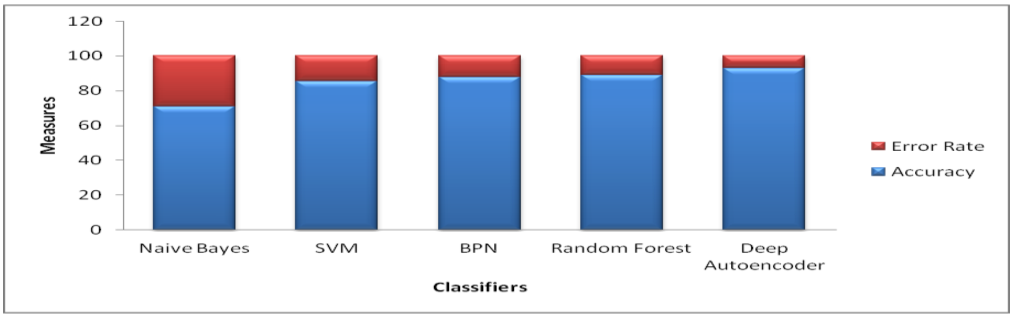

Figure 4. Performance analysis of various classifiers based on accuracy and error rate

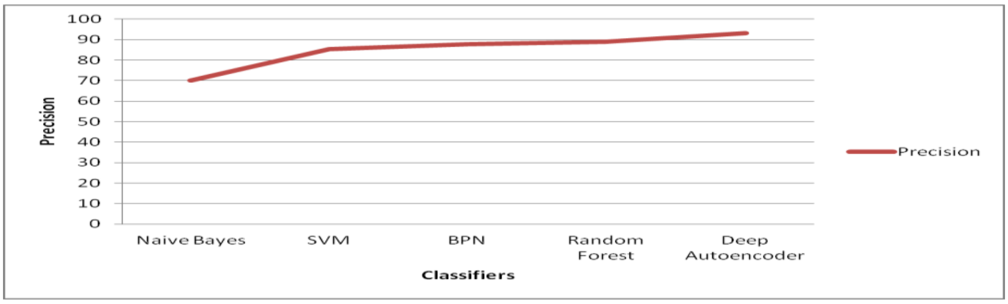

Figure 5. Performance analysis of various classifiers based on precision

The graphical representation of IBM HR employee attrition prediction dataset is presented in figure 4. The quantitative results of the proposed RNN have been compared with standard pattern recognition algorithms and are given in table 3.The measures such as precision, recall, f-measure, accuracy and error rate are utilized to validate the classifiers discussed. From table 3 and from figure 4 - 5, it is clearly understood that the proposed deep autoencoder outperforms the existing classifiers. It is observed that the highest and lowest classification accuracies are 93.10% and 70.10%. The classifiers such as Naïve Bayes, SVM, BPN, and Random Forest exhibits accuracies of 71.14%, 85.47%, 87.86%, and 89.00% respectively. Similarly, the error rates of Naïve Bayes, SVM, BPN, and Random Forest are 28.86%, 14.53%, 12.14%, and 11.00% respectively. The deep autoencoder reveals the precision of 93.10%, recall of 91.15% and f-measure of 92.11%. In summary, the proposed deep autoencoder exhibits more accuracy of 93.10% and less error rate of 6.9% for employee attrition prediction.

5. CONCLUSION

Deep Learning is taking over the world; in day to day life or high-end innovations. In contrast to the conventional machine learning methods, Deep Learning has gained significant attention in the field of artificial intelligence recently. In this study, deep autoencoder is proposed to predict the employee attrition. Deep autoencoder produced high accuracy when compared with the benchmark techniques such as Naïve Bayes, SVM, BPN, and Random Forest. To evaluate the performance of the proposed algorithm several experiments have been conducted on IBM HR dataset. The acquired results expound that the proposed method reveals better results in terms of precision, recall, f-measure, accuracy and error rate. The constructed model enables the HR manager to analyze the past and current employees data to predict the future churners and learn the causes of employee turnover. In future, the hyperparameters of deep autoencoder will be optimized and other deep learning models will be explored for attrition prediction.

REFERENCES

Archita Banerjee, R. K. (2017). A Study on the Factors Influencing the Rate of Attrition in IT Sector: Based on Indian Scenario. Pacific Business Review International , 9 (7), 01.

Barney, J. B., & Wright, P. M. (1998). On becoming a strategic partner: The role of human resources in gaining competitive advantage. Human Resource Management , 37 (1), 31-46.

Bartlett, C. A., & Ghoshal, S. (1997). The myth of the generic managers: new per-sonal competencies for new management roles. California Management Review , 40 (1), 92-116.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives.

IEEE Transactions on Pattern Analysis and Machine Intelligence , 35 (8), 1798-1828.

Booth, S., & Hamer, K. (2007). Labour turnover in the retail industry: Predicting the role of individual, organisational and environmental factors labour turnover in the re-tail industry. International Journal of Retail & Distribution Management , 35 (4), 289-307.

Boushey, H., & Glynn, S. J. (2012). Center for American Progress.

Boxall, P., Purcell, J., & Wright, P. (2007). Human resource management: scope, analysis, and significance.

Oxford Handbook of Human Resource Management,. Breiman, L. (2001). Random forests. 45 (1), 5-32.

Frye, A., Boomhower, C., Smith, M., Vitovsky, L., & Fabricant, S. (2018). Employee Attrition: What Makes an Employee Quit?. SMU Data Science Review , 1 (1), 9.

Golovko, V., Kroshchanka, A., & Treadwell, D. (2016). The nature of unsupervised learning in deep neural networks: A new understanding and novel approach. Optical Memory and Neural Networks , 25 (3), 127–141.

Haykin, S. (2005). Neural Networks. Second edition by Prentice Hall of India.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. Expert Systems with Applications , 39 (1), 1414–1425.

IBM Attrition dataset. ,https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/.

Labov, W., Ash, S., & Boberg, C. (2008). The atlas of North American English: Phonetics, phonology and sound change. Walter de Gruyter.

Leung, K. M. (2007). Naive bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering.

Pfeffer. (1994). Competitive advantage through people: Unleashing the power of the work force. Boston: Harvard Business School Press.

Phillips, P. J. (1999). Support vector machines applied to face recognition. In Advances in Neural Information Processing Systems , (pp. 803-809)

Rifai, S., Vincent, P., Muller, X., Glorot, X., & Bengio, Y. (2011). Contractive auto-encoders: explicit invariance during feature extraction. Proceedings of ICML, (pp. 833–840).

Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. Expert Systems with Applications, , 38(3).

Singh, R., & Om, H. (2017). Newborn face recognition using deep convolutional neural network. Multimedia Tools and Applications , 1-11.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks.Information Processing & Management , 45 (4), 427-437.

Swersky, K., Ranzato, M., Buchman, D., Marlin, B., & de Freitas, N. (2011). On score matching for energy based models: Generalizing autoencoders and simplifying deep learning. International Conference on Machine Learning ICML-11, (p. 1201).

Vincent, P. (2011). A connection between score matching and denoising autoencoders. Neural Computation , 23(7), 1661-1674.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. (pp. 1096-1103). In Proceedings of the 25th international conference on Machine learning .

Wang, G. (2016). A perspective on deep imaging . IEEE Access 4 , 8914–8924.

Yao, Y., Marcialis, G. L., Pontil, M., Frasconi, P., & Roli, F. (2003). Combining flat and structured representations for fingerprint classification with recursive neural networks and support vector machines . Pattern Recognition , 36 (2), 397-406.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Advances in neural information processing systems , 3320–3328.