

# A new method for employee promotion detection based on the hybrid ANN-SVM model

**Shanshan Gong**

Graduate School of Arts and Science, New York University, New York, NY 10003,  
United States

shanshan.gong@nyu.edu

**Abstract.** A company's ability to achieve its ultimate long-term goal highly depends on the potential and motives of its employees. The Human Resource Management team must be able to accurately evaluate a worker's ability in quality work and incentivize him/her using promotion. While it has been quite difficult to measure an employee's aptness for promotion, previously, machine learning classification methods has been researched upon. The aim of this study is to analyze a newly proposed hybrid classification model in detecting potential employee for promotion. The hybrid model combines the strength in feature selection of Artificial Neural Network model and the binary classification power of Support Vector Machine. The model is tested and trained after its hyperparameters are tuned. Afterward, the model is evaluated on the accuracy, precision, recall and F1 score, and compared to the individual SVM and ANN models for performance. The predictions of this proposed model have proven to exceed those two models with a range of 0.15% to 7.26% increase in performances. Note, the precision of ANN model is somewhat higher than the hybrid model. In summary, the hybrid ANN-SVM model has proven to work effectively with both imbalanced and balanced datasets.

**Keywords:** ANN-SVM Method, Employee Promotion Detection.

## 1. Introduction

As companies strive to survive in the face of multinational competitions and to adapt to the new challenges and rapidly changing environments, they must continue looking for innovative technologies and strategies to improve their services and products while reducing expenses – with the ultimate goal of a successful business. Importantly, a business' success builds on its employees, hence, for a company to operate in the long run, its Human Resource Management (HRM) must be able to identify their employees' strength and incentivize them to boost productivity. Promotion has been used to motivate these employees, who in turn will generate quality work. Qualification for promotion might not be quantifiable let alone predictable; thus, researchers have suggested to employ machine learning algorithms to effectively evaluate the employees' aptness for promotion. The ability to accurately predict an employee's promotion would save HRM time and energy in training unsuitable ones, and most importantly it saves the company's budget.

Previously, researches have shed light on various possible ways of predicting an employee's promotion. Most of the papers touch upon multiple classification algorithms with a focus on feature selection methods.

In Liu et al.'s research on employee promotion prediction, they've focused on the role of an employee's position within an organization. After implementing three classification models – Logistic Regression, Random Forest, and AdaBoost – they computed four metrics – accuracy, AUC, Recall and Precision. These metrics allowed them to conclude that Random Forest has the best performance while maintaining acceptable efficiency [1].

Building on top of Liu's research, Long et al. focused on two different sets of features of the employees – “personal basic features” (ex. demographics) and “post features” (ex. departments). By performing k-fold cross validation and hyperparameter tuning on three extra models – K Nearest Neighbor, Support Vector Classifier, and Decision Tree, they've established that random forest model provides the best result, and the post features has a greater overall impact [2].

Wang et al. focuses their research on an innovative feature selection method. By combining Affinity Propagation and SVM sensitivity analysis along with forward selection and backward elimination for selecting the optimized features, they were able to reduce the number of features while increase the model's performance [3].

While this past research explored classifications models in some detail, there is a lack of neural network models. Provided the nature of complicated relationships of features, a neural network model would be applicable in capturing these details. Moreover, due to the importance of effectively identifying the employees for promotion, a combined model that picks up the advantage of the neural network model and maintains the benefit of binary classification model is optimal.

As a company continues to expand, its employee base grows, hence diversifying the talents and skills among the workers. While each individual brings his/her own unique values towards the company, the HRM plays an important role in identify the employees with talents that meets the company's vision. Being able to accurately recognize the eligible candidates for promotion would expedite the promotion cycle cost effectively. Hence, in this research a hybrid model is proposed for predicting whether an employee will be promoted. A particular hidden layer of the previously trained Artificial Neural Network (ANN) is selected and fitted into the Support Vector Machines (SVM) classification model. This novel model combines the ability of neural network models in higher dimension feature selection along with the binary classification strength of classification model. Emphasizing the advantage of both models, the proposed hybrid model is able to obtain an accuracy of 94.09% and a precision of 97.26%.

The rest of this paper comprises of the following sections. In Section 2, a detailed analysis of the data set is provided along with a discussion of the approach. Section 3 addresses the methodologies implemented in predicting the potential employee for promotion and evaluation metrics performed to measure these methods. Results, discussions and significant discoveries are provided in Section 4. Eventually, Section 5 will present a brief conclusion of the project and suggestions for future work.

## **2. Method**

### *2.1. Data description and preprocessing*

The dataset used in this paper comes from Kaggle [4]. It provides various demographic and performance-related information on employees of a large multinational company for the manager position and the positions below manager. The dataset contains 14 different attributes for each employee and there is a total of 54,808 employees. Particularly, one of the attributes is “is\_promoted” and it has two possibilities – 0 if the employee is not promoted, 1 if the employee is promoted. Naturally, this attribute becomes the y variable given the ultimate goal of the research in detecting potential employee for promotion, while all other 13 attributes are regarded as X features. The descriptions of the features are provided in Table 1.

**Table 1.** Feature descriptions of the collected dataset.

Feature	Description	Data Type	Remark
<b>employee_id</b>	Unique ID for employee	Continuous	Example: 13614
<b>department</b>	Department of employee	Categorical	Example: Technology, Operations
<b>region</b>	Region of employment	Categorical	Example: region_1, region_3
<b>education</b>	Education level of employee	Categorical	Example: Master's & above
<b>gender</b>	Female or male	Binary	f = female, m = male
<b>recruitment_channel</b>	Employee's recruitment channel	Categorical	sourcing, referred, other
<b>no_of_trainings</b>	Number of other trainings <sup>a</sup>	Categorical	1, 2, ..., 10
<b>age</b>	Age of employee	Continuous	Range: 20 ~ 60
<b>previous_year_rating</b>	Employee's previous year rating	Categorical	1, 2, 3, 4, 5
<b>length_of_service</b>	Length of service in years	Continuous	Range: 1 ~ 37
<b>KPIs_met &gt;80%</b>	If KPI is met more than 80%	Binary	1 = met, 0 = not met
<b>awards_won?</b>	If awards won in previous year	Binary	1 = won, 0 = didn't win
<b>avg_training_score</b>	Average training score	Continuous	Range: 39 ~ 99
<b>is_promoted</b>	Recommended for promotion	Binary	1 = promoted, 0 = not promoted

<sup>a</sup> Trainings on skills such as soft skills, technical skills, etc. that are completed in previous year.

A few pre-processing steps are performed beforehand to ensure the performance of models in the next step. Firstly, due to the uniqueness of column “employee\_id” for each employee and its lack of trend, this column is dropped. Noting the repetitive word in the “region” feature – region\_1, region\_2, ... – a numerical version is created – 1, 2, .... Further, acknowledging the categorical nature of various features and the lack of ordinal relationship between the values, one-hot encoding technique is performed on “department”, “education”, “gender”, and “recruitment\_channel” attributes. Moreover, since the rows with missing value consist of a relatively small portion of the overall data, these rows are dropped, resulting in still over 50 thousand rows of feasible data. In addition, the data is normalized using the MinMaxScaler to refrain the model from picking up the different scales used for each feature as a pattern, thus create bias in the result. Lastly, the pre-processed data is split into 70% for training models and 30% for testing their performances.

## 2.2. Models

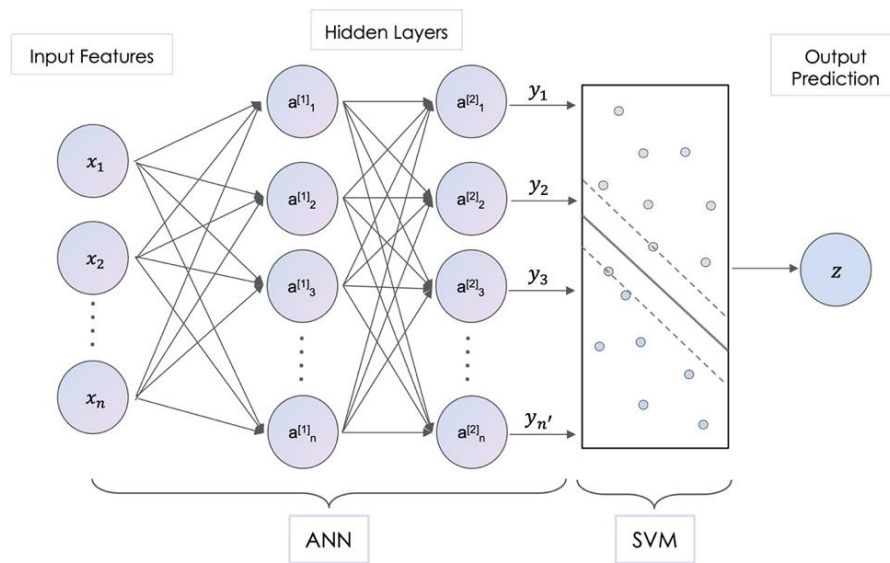
This paper focuses on the state-of-art neural network and classification models to solve the binary classification problem. ANN and SVM models are implemented and compared with the hybrid ANN-SVM model after tuning their hyperparameters.

**2.2.1. ANN.** Artificial Neural Network is a learning algorithm that was designed to mimic the structure of neurons in human brains. The model consists of three types of layers: an input layer for the original data, one or more hidden layers are applied to the activation function to transform the data, and finally an output layer for the predictions. While training the model, gradient descent method is applied on the predicted and actual output through back propagation to achieve the desired outputs. While traditional machine learning algorithms tends to have limitations in adequately fitting the data, ANN's highly interconnected nodes grant the model with the ability to capture complex relationships and trends [5-7].

**2.2.2. SVM.** A Support Vector Machine is a supervised machine learning algorithm that focuses on binary classification problems. The model has the advantage of performing efficiently under a limited amount of data [8, 9]. A decision boundary – also known as the hyperplane – is created between the two classes of data, which is placed as far as possible from the closest points of the two classes. This linear

classifier model can also be used to model higher dimensional, non-linear data through specifically setting its kernel hyperparameter [10].

**2.2.3. Hybrid ANN-SVM.** The proposed hybrid model combines the ANN model with SVM classifier to solve complex binary classification problems. As shown in Figure 1, the ANN model runs to extract crucial features from the original dataset, then these modified features will be fed into the SVM model for classification prediction. As mentioned above, ANN has an advantage in learning complex and arbitrary relationships between inputs and outputs. Hence, it would be used to analyze complicated relationships through its hidden layers – performing abstract feature selection at a higher dimension. Furthermore, since SVM has the ability to efficiently and effectively perform binary classification tasks, it is used to perform classification on the selected features of a certain layer of ANN. Binary classification performed upon these features would result in a better performance than the previously proposed methods [11].



**Figure 1.** Hybrid ANN+SVM model architecture diagram.

### 2.3. Implementation details

Firstly, Scikit-learn and TensorFlow libraries are imported for successfully implementing the two models. Hyperparameter tuning is executed to ensure an optimized performance of metrics including accuracy, precision, recall, and F1. Specifically, the ANN model has three hidden layer: 70 nodes in its first hidden layer and uses the “relu” activation function; 40 nodes in its second hidden layer with “sigmoid”; and 10 nodes in its third hidden layer with “relu”. The output layer has 1 node and uses “sigmoid”. Noted that the ANN model is trained under the “Adam” optimizer, “binary\_crossentropy” loss function, 0.0001 “learning\_rate” and “accuracy” evaluation metrics. In the SVM model and the hybrid model, on the other hand, the “kernel”, “c”, “degree”, and “layer” hyperparameters are tuned to achieve each model’s best performances.

Additionally, since the data is highly imbalanced, the recall and F1 score are quite low, however, since the goal of the project is to achieve a relatively higher score using the hybrid model, no additional steps are performed on the data to balance out the classes. Also, the pre-processing, training, and testing stages are all completed on the CPU.

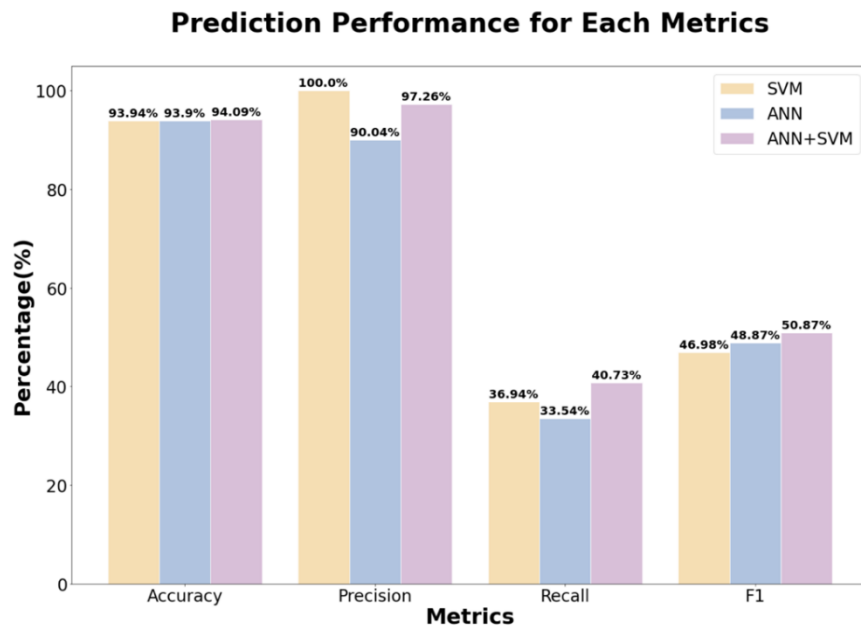
### 3. Results and discussions

#### 3.1. Results

By tuning the hyperparameters mentioned above, the best performances reached by each of the individual models are recorded. Table 2 displays the testing data performances of the three different models. For a more visualized representation, Figure 2 graphs the predictions of the three models for each of the four metrics used. Thus, as seen in the graph, the hybrid ANN+SVM model is superior to the ANN model since the performance of each of the four metrics – accuracy, precision, recall, and F1 score – all outperforms those of the ANN model. Moreover, comparing the hybrid model to SVM, there is also an increase in the performance of accuracy, recall, and F1 score. Note that the precision score of the hybrid model is not ideal in comparison to the surprisingly high precision score of SVM (100%). However, despite this minor shortcoming, as seen in Table 2, the hybrid model shows a range of 0.15% to 7.26% increase in performances in comparison to both ANN and SVM models alone.

**Table 2.** Feature descriptions of the collected dataset.

	ANN	SVM	Hybrid ANN+SVM
<b>Accuracy Score</b>	93.9037%	93.9432%	94.0879%
<b>Precision Score</b>	90.0407%	100.000%	97.2637%
<b>Recall Score</b>	33.5352%	36.9417%	40.7267%
<b>F1 Score</b>	48.8693%	46.9775%	50.8682%



**Figure 2.** Metrics comparison of performance.

#### 3.2. Discussion

Considering the established properties of the ANN and SVM models, the outperformance of the hybrid ANN+SVM model in various metrics becomes more comprehensible. The hybrid model effectively combined the advantages of the two individual models. Specifically, ANN models often excel in their abilities to extract characteristics that are more abstract and more complicated from the original data set. In comparison to simply using classification models where several features of the original data might not be enough for a classification model to categorize accurately and precisely, the hidden layers in

neural network models generates additional nodes which in turn increases the number of dimensions, hence providing the model with additional features to work with during classification. Moreover, these extracted features are more essential for the output than the original features due to the mechanism in feature selection of ANN models. In addition, since SVM models performs well for binary classification problems, thus feeding these features into this effective binary classification model would result in better performances.

Furthermore, based on the higher values achieved in various metrics analysis – precision, recall, and F1 scores, the hybrid ANN-SVM model has proven to work well with imbalanced datasets. Importantly, the improved accuracy performance in comparison to ANN and SVM models alone, also provides evidence for the hybrid model to exceed under balanced data.

#### 4. Conclusion

In summary, this paper aims to identify an employee's strength and incentivize them to boost productivity by investigate various possible ways of predicting an employee's promotion. Specifically, a hybrid ANN+SVM model is designed and constructed by incorporating the properties of these two models to improve the performance of detecting employee for promotion. The accuracy, precision, recall and F1 score are evaluated for the model and compared to the scores of ANN and SVM models alone to determine its effectiveness. According to the analysis, the hybrid model is able to achieve a more ideal performance in forecasting the employees for promotion than these two individual models under various hyperparameter settings. The proposed hybrid ANN+SVM model is a practical and effective method in detecting potential employee for promotion. Thence saving time and budget for the company. An area of possible future research from this study is to develop a hybrid model for image classification, where rather than extracting features from a layer of the ANN model, a convolutional neural network model would be more suitable for computer vision research.

#### References

- [1] Liu J et al. 2019 A data-driven analysis of employee promotion: the role of the position of organization IEEE Int. Conf. on systems, man and cybernetics (SMC) (Bari) p 4056–62
- [2] Long Y et al. 2018 Prediction of employee promotion based on personal basic features and post features Proc. of the Int. Conf. on Data Processing and Applications (Guangzhou) p 5–10
- [3] Wang Q Li B and Hu J 2009 Feature selection for human resource selection based on affinity propagation and SVM sensitivity analysis World Congress on Nature & Biologically Inspired Computing (NaBIC) (Coimbatore) p 31–36
- [4] Kaggle 2020 Hranalysis <https://www.kaggle.com/datasets/shivan118/hranalysis>
- [5] Otchere DA Ganat TO Gholami R and Ridha S 2021 Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models Journal of Petroleum Science and Engineering vol 200 p 108182.
- [6] Yu B Wang YT Yao JB Wang JY 2016 A comparison of the performance of ANN and SVM for the prediction of traffic accident duration Neural Network World vol 26 number 3 (Barcelona) p 271
- [7] Yu Q et al. 2020 Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV China Communications 17(3) p. 46-57
- [8] Foody GM and Mathur A 2004 Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification Remote Sensing of Environment vol 93 number 1–2 (Amsterdam: Elsevier) p 107–17
- [9] Emmanuel-Okereke IL and Anigbogu SO 2022 KNN and SVM Machine learning to predict staff due for promotions and training
- [10] Huang S et al. 2018 Applications of support vector machine (SVM) learning in cancer genomics Cancer Genomics & Proteomics vol 15 Issue 1 p 41-51
- [11] Shylaja S and Muralidharan R 2019 Hybrid SVM-ann classifier is used for heart disease prediction system IJEDR vol 7 number 3 p 2321–9939