iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XI **Month of publication:** November 2022

DOI: https://doi.org/10.22214/ijraset.2022.47376

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

Human Resource Management by Machine Learning Algorithms

Mohammad Tehaseen Hussain[1], Mohammed Abrar Baig[2]

Abstract: The main object of Human Resource Management using machine learning is to analyze employees' interest in their respective companies and make predictions about who will stay or leave the organization. The data is extracted using data scraping techniques and stored in CSV format. Data collected through this technique contains different features, and with the help of ML algorithms, it leads to predictions. This analysis helps the manager to make conclusions about who will stay or leave the organization; with this, the manager can approach a way to let stay a worthy employee in the organization.

In our study, we used different techniques like feature scaling and SMOTE. According to the results, the suggested methods like random forest and XG boost classifier. The accuracy rate(%) values for the outcomes produced by the suggested methods will lead us to the conclusion.

Keywords: Support Vector Machine (SVM), K- nearest neighbor (KNN), Artificial neural network(ANN), Synthetic minority oversampling technique(SMOTE), Regression, Hr Analytics, Classification, Random forest Regressor, Convolution Neural Network(CNN).

I. INTRODUCTION

HR professionals play an important role in the organization, planning, coordinating, supervising strategic planning, controlling the hiring process, maintaining legal compliance, and Business planning. This helps in decision-making and integrates HR with Business goals. In this Machine Learning model, we use scraped data to predict whether the employee is staying or leaving the organization. Most organizations can use this method and predict the outcome. We have used different algorithms to get the best accuracy model. The algorithms used for processing are

1) Support Vector Machine(SVM)
2) K-Nearest neighbor(KNN)
3) Random Forest
4) XGBoost Classifier
5) Artificial Neural Network

The data consists of different aspects of the employee like their salary, hours working on the project, Productivity, operation management, Hiring Process, Number of employees, Financial reports and Taxes, etc. the data collected is unstructured data. It is changed into structured data and is stored in CSV format. The next step is preprocessing in order to remove outliers and replace null values with mean, median, or mode. Once preprocessing steps are completed, Machine learning algorithms are applied to the data to make predictions or classify the data.

II. HISTORY & BACKGROUND

Machine Learning is the subset of Artificial Intelligence that uses data to make predictions and recommend better results. Suppose if an organization wants to Improve its profits or productivity of the employees they can use a recommendation system to get a better understanding of the problem and to improve efficiency of the employees. Generally, ML was introduced in the 19th century, but it is not implemented due to little computational power. Moreover, recommendation systems suggest most pertinent items to users and filter the data using various ML Techniques. It records the users preferences and inclinations and then processes alternatives that are consistent with those preferences.

A. Algorithm Used

1) *Support Vector Machine(SVM):* It is one of the most popular Supervised Learning algorithms, including classification and regression analysis. The main aim is to create the best fit line to differentiate the data points in a category. Whenever a new data point is given, it is classified based on the training data.

2) *K-Nearest neighbor(KNN):* In machine learning, KNN is one of the simplest algorithms. It stores the available data and puts it in the category when new data comes. It is also known as the lazy learner algorithm. It can also be used in classification and well as regression problems.
3) *Random Forest:* Random forest is popular machine learning that can solve both classification and regression problems. It contains an N number of decision trees and divides the dataset into subsets which leads to higher accuracy and prevents overfitting problems.
4) *XGBoost Classifier:* A gradient boosting framework is used by the ensemble machine learning method called XGBoost, which is based on the decision tree. Artificial neural networks frequently outperform all other algorithms or frameworks in prediction problems requiring unstructured knowledge (pictures, text, etc.). However, decision tree-based algorithms are considered the best when small to medium amounts of structured/tabular data are involved.
5) *Artificial neural network(ANN):* An artificial neural network is usually a computational network based on a biological neural network that refers to the human brain. It contains three layers: the input, hidden, and output. Hidden layers perform all the calculations to find hidden features and patterns in the data set

III. PROPOSED ALGORITHMS

Data is collected through scraping techniques with different labels and has been preprocessed using different techniques to remove outliers, encoding, feature scaling, etc. Though the system flow diagram for the proposed system is shown in Fig. 1. It involves the collection of data through techniques like data scraping, and the collected data is preprocessed, encoded using trained datasets, testing the algorithm, and predicted the results, which are shown in the system flow diagram. The labeled data is encoded for better prediction

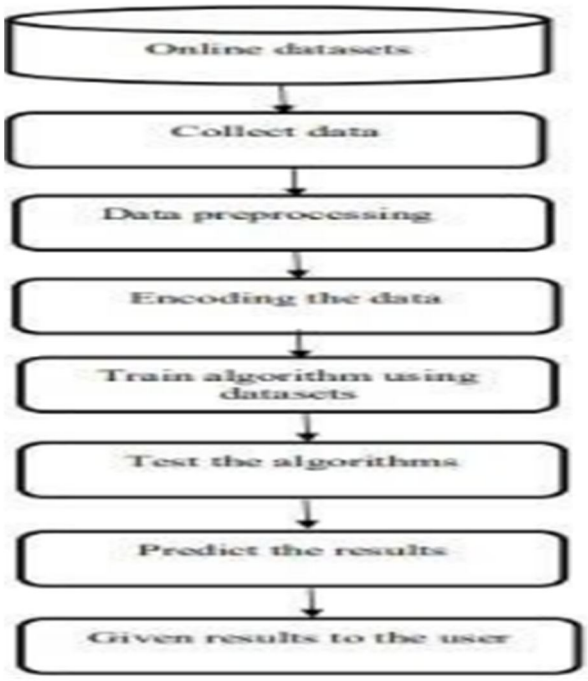


Fig 1.1 System flow diagram for the proposed system

A. *Data Preprocessing*

Preprocessing is one of the important steps in Data mining. After extraction, Preprocessing was done to clean the data by the library in python. Cleaning the data involves detecting the null and duplicate values, and if any are found will be replaced by mean or medium. Outliers are known as noise in the dataset, which has been detected and replaced by medium imputation. Dropping some of the columns in the dataset because they are not useful in predicting the target variable. Ordinal Encoding helps in the classification process, which converts data into binary format, i.e., 0 and 1. For accurate results, the data has been cleaned with accurate features. The data is imbalanced, so we used one of the most popular oversampling techniques, SMOTE, to handle imbalanced data.

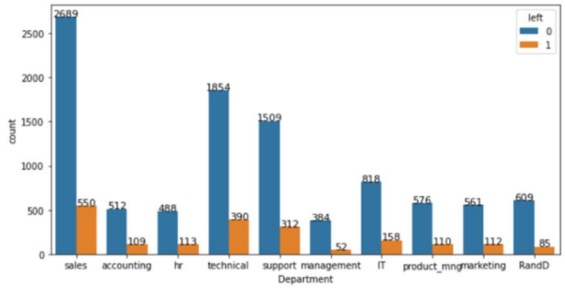


Fig 1.2 Left or stayed in different departments

SMOTE combines already existing minority instances to create new minority instances. Split the data in an 80:20 ratio for training and testing the data. Feature scaling is a technique to scale the data to handle weight greater values, bringing all the values in one scale.

B. Visualization

The accuracy values obtained from each algorithm have been considered, and the best model has been chosen. The metrics such as precision, recall, F1 score, and confusion matrix are taken into consideration when choosing the best model.

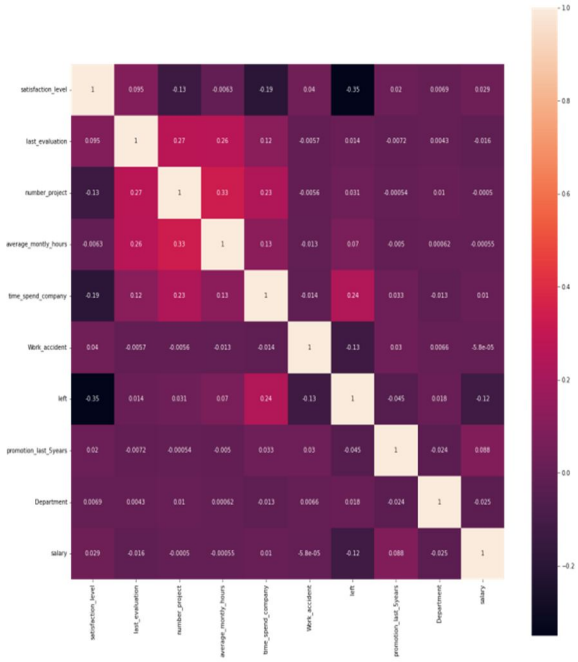


Fig 1.3 Heat map

$$Precision = \frac{TP}{TP + FP}$$

\

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

IV. EXPERIMENT AND RESULT

The experimental setup was completed in Windows, Anaconda environment, and python version 3.9.3. 5 different classifiers are trained, and predictions are made. Table I displays the accuracy and precision of different classifiers. SVM accuracy was 70%, KNN accuracy was 95%, Random forest accuracy was 98%, XG boost accuracy was 98%, and ANN accuracy was 95%. Random forest and XG boost provide very good accuracy. When compared with other models.

TABLE1 Model accuracy in training dataset

Algorithms	Accuracy	Precision	Recall
SVM	70%	32%	17%
KNN	89%	63%	89%
Random Forest	97%	96%	90%
XG Boost	98%	98%	98%
ANN	95%	92%	89%

TABEL2 Model accuracy in testing dataset

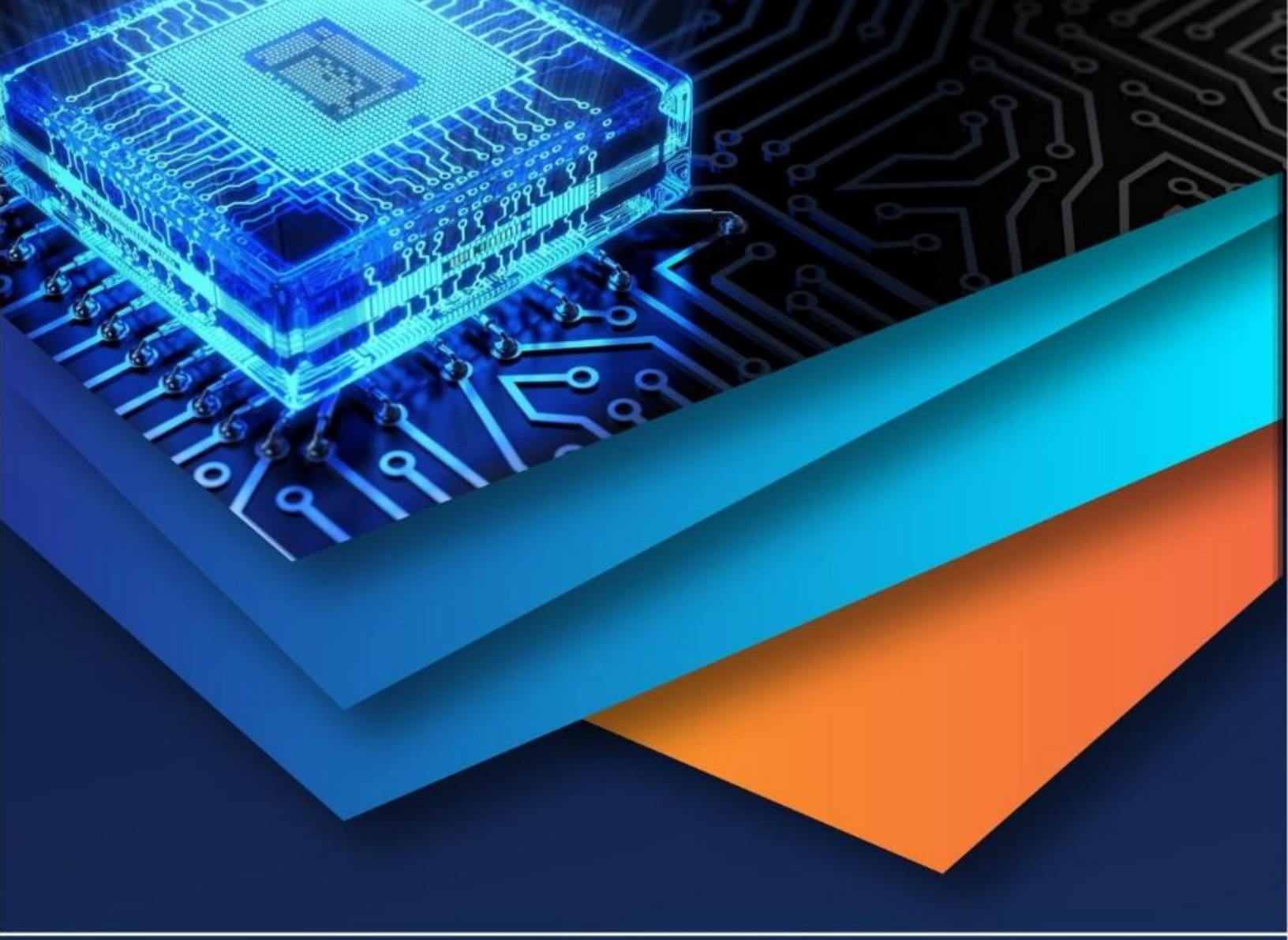
Algorithm	Accuracy	Precision	Recall
SVM	71%	70%	72%
KNN	95%	92%	98%
Random Forest	98%	98%	98%
XG Boost	98%	98%	98%
ANN	95%	94%	92%

V. CONCLUSION

The score is calculated and plotted in a graph using different features like employment, projects, taxes, management, etc. by using different features like employment, projects, taxes management, etc. Moreover, the accuracy acquired is 91%, but by using Hyperparameter Tuning, the model's accuracy is improved; hence we got our best model by using a random forest classifier with an accuracy of 98%. Furthermore, by applying algorithms like SVM K-Nearest neighbor(KNN), Random Forest, XGBoost Classifier got the accuracies of 98.5%, 99%, and 97%, and it is concluded that Random Forest and XG Boost work best with this dataset.

REFERENCES

[1] Hr Analytics: Connecting Data and Theory by Rama Shankar
[2] Hr analytics practical approach using python by Bharti Motwani
[3] Data Science and analytics by V.K Jain
[4] Data Science and Machine Learning by N. Meenakshi and Rajakumar
[5] Hr Analytics Essentials you always wanted to know (self-Learning Management series) by Vibrant Publishers
[6] Hr analytics: Quantifying the Intangibles by Anshul Saxena
[7] Applying Advanced Analytics to Hr Management Decisions: Methods for Selection, Developing Incentives and Improving Collaboration by James C Sesil, Pearson Education



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)