MACHINE LEARNING TO IMPROVE PRODUCTIVITY EMPLOYEES RATING IN SOFTWARE COMPANY

Manoj Kumar Sarakanam
dept. of Computer Science
University of Central Missouri
Missouri, USA
mxs19150@ucmo.edu

Devi SatyaSri Gundumogula
dept. of Computer Science
University of Central Missouri
Missouri, USA
dxg20850@ucmo.edu

Nutulapati Praveen Kumar
dept. of Computer Science
University of Central Missouri
Missouri, USA
pxn20480@ucmo.edu

Venkata Malakonda Prudhvi Tanikonda
dept. of Computer Science
University of Central Missouri
Missouri, USA
vxt26300@ucmo.edu

***Abstract*— Since employee performance has a detrimental impact on operational productivity and long-term evolution objectives, it has been recognized as a crucial issue for businesses. In order to anticipate workplace efficiency, businesses apply machine learning algorithms to address this issue. Organizations can take action on employee succession planning or employee preservation thanks to precise forecasts. Although it is typically smaller in comparison to other areas of the company's information systems and is obviously related to its objectives, the data for the modeling problem comes from HR Information Systems. Repetitive values in the data make predictive models susceptible to over-fitting and thus unreliable. This leads to the occurrence of redundant values in the data. we applied Support Vector Machine (SVM) and Random Forest Classifier (RFC) to this dataset, which allowed us to predict the attrition rate with an accuracy of 86.0%. The F1-score and accuracy, two key classification metrics, have been used to express the results.**

Keywords—Employee Performance,Behaviour,Random Forest Algorithm(RFA).

I. INTRODUCTION

Employee performance concerns have become more important in businesses because of their negative effects on anything from office morals and productivity to project continuity and long-term growth goals. Businesses can address this issue by employing machine learning techniques to predict the likelihood of losing employees. By doing so, officials and human resources will have the foresight necessary to take prompt preservation or strategic planning action. Most businesses did not place a high priority on making investments in effective human resource programs that would gather information from workers while they were at work. One of the important elements is the lack of experience with costs and benefits. HRIS's return on investment is currently challenging to quantify. Data redundancy is the outcome, that limits how broadly these techniques can be applied.

Employee performance is a problem, and the major machine learning techniques are used to fix it. The goal of this study is to demonstrate how gradient boosting can be used to improve the presented algorithms, particularly in terms of the generalizability and prevalence of redundant data. These are accomplished by categorizing the weakening problem as a classification problem, designing it using supervised methods, and employing HRIS data from a global corporation. By comparing the classifier's higher accuracy with other approaches and explaining why they perform better, the conclusion is made.

Software challenge managers compare the actual final touch of activities in opposition to the development reviews filled by way of project contributors to identify substantial deviations from the estimated schedules and manage software program assignment risks. However, quantitative measurements are restricted due to the layout of mission documents, which are mostly herbal languages. The proposed approach can recognize entailment from textual content successfully and outperform other textual entailment approaches. The textual entailment approach to venture monitoring and manipulation, which not only reduces the assignment value and human effort however also provides a foundation for challenge managers to qualitatively evaluate the overall performance of every challenge member. The employee rating in the company is taken and analyzed with a different type of attributes. The

GIT HUB LINK:

employee ratings are provided based on the project reviews, client feedback and client satisfaction level about the project.

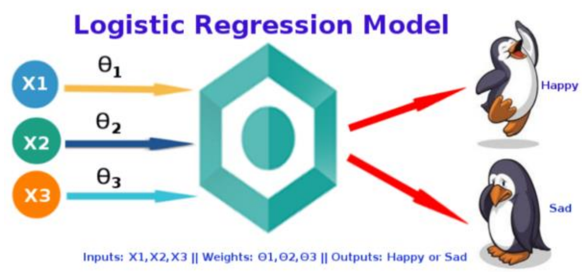
The project evaluation can be tested with the machine learning algorithm prediction results.

Since the random forest algorithm will be used for analysis , the accuracy of the algorithm result will be helpful to evaluate the results. The accuracy score of the algorithm in the project helps to evaluate the dataset.

High performance has several negative effects on a company. It is challenging to replace employees who possess specific skill sets or are industry experts. It has an impact on present staff productivity and ongoing research. Finding replacement workers comes with its own costs, such as those associated with coaching and recruitment. On the other side, new hires could encounter learning curves before they have the same amount of company or technological experience as seasoned internal staff.

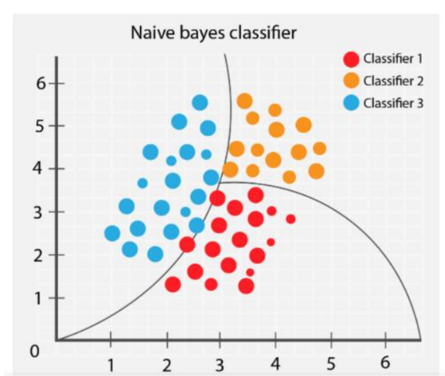
A. *Logistic regression*

One of the most common linear models for classification is the maximum entropy/logistic regression approach. A frequent type of regression used for conditional or categorical predictions of variables is logistic regression. To prevent overfitting, it is frequently employed in conjunction with regularization in the form of L1- or L2-norm based penalties. An L2-regularized logistic regression was used for this study. This technique gets the posterior probabilities and calculates the model's parameters by assuming a model for the situation.



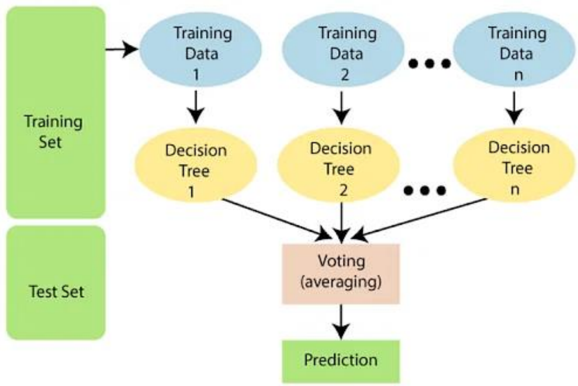
B. *Naive Bayesian*

Because of its simplicity and effectiveness, Nave Bayes is a popular categorization technique. Naive Bayes uses the assumption that all variables are conditionally independent to classify data according to the likelihood of arrival. To approximate the parameters, such as the means and variances required for classification variables, the classifier only requires a modest amount of training data. Additionally, it manages real, discrete data.



C. *Random Forest*

The tree-based Random Forest algorithm is a well-liked ensemble learning technique. The 'assembly' type here is to inflate. In bagging, each tree is generated independently using a particular data set bootstrap sample, therefore subsequent trees are not dependent on preceding trees. The final decision on the forecast is made by a simple majority vote. Standard trees and random forests differ in how each point is divided using the latter's optimal split of all variables. Every node at that node in a random forest is segregated using the maximum among a subgroup of predictors that were chosen at random. It has great resistance to overfitting thanks to the additional random layer.



D. K-Nearest Neighbor(KNN)

The labeling of datasets based on the evaluation of their nearest neighbors is known as nearest neighbor classification. The method is more frequently known as the k- Nearest Neighbor (k-NN) classification because it is advantageous to discover several neighbors.

E. linear discriminant analysis

To optimize the variation across the categories in proportion to the category variance, discriminant analysis necessitates the creation of one or more discriminating functions[14].
A linear combination of two or more independent variables that successfully distinguishes between two or more different categories or groups is referred to as a linear discriminant analysis.
The likelihood of a class belonging to a specific member or observation is then determined using the z-scores generated by the distinguishing methods.
Another important thing to keep in mind when using LDA is that the functions should either be continuous or linear in form.

II. MOTIVATION

When configured, the program automatically keeps track of any data trails that a company might make available for each of its employees. The system uses machine learning techniques to learn the typical workflow for various workers. Each employee receives a "productivity score" from the software, which has identified their typical behavior. The ratings of employees in a company can still be compared even though they perform different activities. An evaluation of productivity also reveals whether employee effort increases or decreases productivity.

The outcomes of the machine learning algorithm predictions can be used to test the project evaluation. The correctness of the algorithm result will be useful to analyze the outcomes because the random forest algorithm will be employed in the analysis. A productivity score also shows whether employee efforts boost or depletes productivity.

Additionally, motivated employees execute their jobs much better than those that lack motivation. Employee aptitude and skill levels greatly increase with improved motivation. As a result, people participate more actively in the production process and produce superior goods.

III. MAIN CONTRIBUTION & OBJECTIVES

- Our objective of the project is the employee rating productivity improvement will be the Python based application which contributes to improving the production of the software company.
- It will be helpful to the software company to improve the skills of the employee based on the project requirements and also to improve the company’s productivity level.
- In this work, the dataset containing the employee rating will be taken into consideration. The preprocessing will be applied to the dataset and the noisy and null value data will be removed from the dataset.
- After the data will be analyzed and visualized for further processing. The machine learning algorithm will be chosen to make the analysis.
- The application will be developed with Google Colab Python TOOL as the project can be directly executed

in any type of computer system with an internet connection.

- There is no need for any specific software to be installed in the user system. The Colab Tool helps to develop and run the application inside the cloud server.

IV. RELATED WORK

The way an employee performs might be viewed as a leak or termination of the employer's intellectual capital[2]. The majority of performance literature divides outcomes into voluntary and involuntary categories. The obligatory performance is the subject of this analysis. The strongest indicators of voluntary performance, according to the meta-analytical analysis of mandated performance studies[3], include age, tenure, income, overall job happiness, and employee perceptions of justice. Other comparable study findings have demonstrated that age, gender, ethnicity, education, and marital status are crucial indicators for determining the success of volunteers. Pay, working conditions, job satisfaction, supervision, promotion, recognition, opportunity for growth, burnout, and other factors are also studied.

An organization will suffer various negative effects from high performance. It might be challenging to replace employees with specific skill sets or industry experts. Current staff productivity and ongoing research are both impacted. Finding replacement workers comes with its costs, such as coaching fees and recruitment expenditures. On the other hand, new hires could encounter learning curves as they catch up to seasoned internal workers in terms of technology or business knowledge.

To give them a perspective for operation, organizations solve this problem by using machine learning approaches to forecast results. The Below Information provides a quick summary of the literature review's conclusions. The subsequent sections of the study will show how the classification models presented are insufficient to deal with the scaled noise in HRIS.

The Summary List of existing works on the specified domain is as follows:

Jantan, Hamdan and Othman studied the Data Mining techniques for performance prediction of employees which deals with theC4.5, decision tree, Random Forest Algorithm, Multilayer Perceptron, and Radial Basic Function Network

Nagadevara, Srinivasan and Valk Studied on the relationship between withdrawal behaviours like lateness and absenteeism, job content,texture and demographics on employee performance and deals with the techniques of Artificial Intelligence and Neural Networks,Logistic Regression, Classification and Regression Tree and discriminant analysis

Hong, Wei and Chen Studied about the Feasibility of applying the Logistic and Probitmodels to employee voluntary Performance which deals with the Logistic regression model and probability model.

Marjorie Laura Kane Sellers Studied the various personal as well as work variables impacting employee voluntary performance using Binomial Logistic Regression.

Alao and Adeyemo Analysing attributes using Decision tree Algorithm

Afef Saidi et al. (2020) [4] presented an innovative audio-based method to detect depression using a hybrid model. Their model combines convolutional neural networks (CNN) and Support Vector Machines (SVM) [13], where SVM is deployed in place of the fully connected layers in CNN. In this proposed model, feature extraction was performed using CNN, and the classification was performed using SVM. They achieved an accuracy of 68% using the hybrid model compared with 58.57% achieved with the CNN model.

Akkapon Wongkoblap et al. (2019) [5] collected Facebook users' data from 2007 to 2012 to build a predictive model for detecting symptoms of depression. They used multiple instances of learning neural networks to create their predictive model. This enabled them to develop their model using a few labeled bags instead of requiring all of the labels of the instances used. They achieved maximum accuracy of 74.51% and a precision of 80% in detecting depressed users based on the content on their social network account

Dilip Singh Sisodia et al. (2017) [6] used the HR analytics dataset sourced from Kaggle and tried to build a model that predicts employee churn rate. A correlation matrix and heatmap were generated to show the relation between the attributes. In the experimental section, a histogram was created to compare left employees vs compensation, department, satisfaction level, etc. They used various machine learning algorithms such as Support Vector Machine (SVM), Random Forest Classifier (RFC) [12], K-Nearest Neighbor (KNN) [14], and Na¨ıve Bayes classifier [15] for prediction purposes. Based on this research, we have incorporated the same algorithms for training and testing our model.

S. S. Alduayj et al. (2018) [7] used a 3 step experiment process to determine the employee attrition rate. In the first experiment, they used an original imbalanced dataset using SVM with various kernel functions KNN, and Random Forest. They concentrated on reducing class imbalance using the adaptive synthetic (ADASYN) approach, retraining the new dataset using the above-mentioned machine learning models. Furthermore, they performed undersampling of the data to achieve a balance between classes. Finally, training an ADASYN-balanced dataset with KNN with K = 3 resulted in the highest performance, with a 0.93 F1 score. They achieved an F!-score of 0.909 while using 12 features out of 29, using Random Forest Classifier and Feature Selection. The essential idea of ADASYN [11] is to use a weighted distribution for different minority class examples according to their level of difficulty in learning, where more synthetic data is generated for minority class examples that are harder to learn compared to those minority examples that are easier to learn.

A carrier company ordered the integration of a revolutionary algorithm named the Data Mining Evolutionary algorithm(DMEL) (2003) [8]. Its prime objective was to anticipate the consumer's attrition rate and the chances of them leaving. It was established that if a consumer is leaving, then a set of loyalty programs, including special offers and discounts, are offered by the company to retain the consumer. Applying the model to real-time data showed accurate results by depicting stimulating rules of classification and distinct attrition rates. Using this concept, our model will give the employer/admin suggestions to retain employees on the verge of quitting.

M. Deshpande et al.(2017) [9] successfully implemented emotional analysis based on Twitter feeds, primarily focusing on depression. the feed was classified as negative or neutral, based on a specially constructed list of words depicting depression tendencies. They adopted a unique approach to conducting their experiments. By implementing Naive-Bayes Classifier and SVM, they achieved a maximum accuracy of 83.0% in predicting depression tendencies.

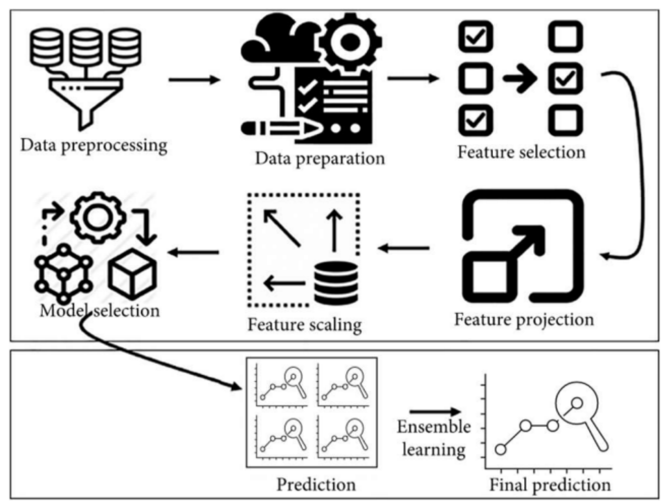
A study conducted by AR Subhani et al. (2017) [10] found that the human brain is the most affected organ while undergoing stress. This study can be applied to learn the changes and stress that a person with mental illnesses like depression, anxiety, etc., goes through. Several features from the signal analysis of the Electroencephalogram (EEG) on the affected person can be extracted. Classification of the extracted features using algorithms like Decision Tree Classifier (DTC) [15], Logistic regression, etc., providing results that were used to differentiate between stress levels, which helped identify psychological disorders.

V. PROPOSED FRAMEWORK

Building a machine learning project involves several technological advancements, but not all of them are required, and everything depends on the information. In this assignment, we'll build a machine learning project that analyzes employee performance and learn about the lifespan of machine learning projects.[2][3] Even though it appears to be simple work or just building a model, the dataset we have is enormous and requires a lot of element designing and preprocessing.

The goal of this research is to promote a commonsense understanding of the positive and bad experiences that employees have as a result of technological stress and consciousness reception created by humans. It separates the performance in the software industry from issues linked to human resource improvement.[4]

Various Phases in Machine Learning



Phase 1: Preprocessing of Data

Before pre-processing, we first collect the data we are interested in collecting to apply classification and regression methods. Put simply, data preprocessing is the act of transforming raw data into a format that can be understood. Real-world data is frequently incomplete, untrustworthy, and likely to contain errors. A tried-and-true method to address these issues is preprocessing data. Data pre-processing is the procedure that gets raw data ready for further processing. We used the standardization strategy to perform pre-processing on the UCI dataset. This step is extremely important because the caliber and volume of data you collect will have a direct impact on how accurate your prediction model may be. We collect samples of both benign and malignant breast cancer in this case. We will use this as our training data.

Phase 2: Data Preparation

Data preparation is the process of importing our data into the appropriate location and preparing it for usage in our machine-learning training. After initially gathering all of our data, the ordering will be randomized.

Phase 3: Feature Selection

In machine learning and statistics, as well as variable selection and attribute selection, the process of selecting a subset of pertinent features to be utilized in the construction of a model is known as feature selection.

Wisconsin Data File and Feature Selection Data on breast cancer diagnosis Out of the 31 parameters included in the Kaggle library, we selected about 8–9 of them. Our aim parameter is the diagnosis of breast cancer, whether it is malignant or benign. The Wrapper Method was used for Feature Selection. Concave points worst, Texture worst, Area se, Area worst, Texture mean, Smoothness worst, Radius mean, Smoothness mean, and Symmetry means are the significant features discovered by the study.

Phase 4: Feature Projection

Using feature projection, data from a high-dimensional space is converted into a lower-dimensional space (with fewer attributes). Both linear and nonlinear reduction procedures may be used, depending on the type of correlations between the features in the dataset.

Phase 5: Feature Scaling

The characteristics in the dataset frequently have a wide range of magnitudes, units, and ranges. The Euclidean distance between two data points is, nevertheless, calculated by the majority of machine learning algorithms.

It is necessary to scale up each attribute to the same magnitude. This can be achieved by scaling.

Phase 6: Model Selection

Before the machine is trained on the data, the input and output of the data are labeled as part of the supervised learning process.

The model can assess fresh data and generate future predictions after being trained on existing data. They are separated into classification and regression strategies. When a regression is the result, the value is real or reliable, such "wage" or "weight." There is a classification challenge when the result is a category, for as classifying emails as "spam" or "not spam." Giving unlabeled or uncategorized data to the computer and allowing the algorithm evaluate it without being given any instructions is known as unsupervised learning.

Phase 7: Prediction

Machine learning uses data to generate solutions to questions. As a result, it is during the prediction or inference stage that we are able to respond to numerous questions. This is the culmination of all of our efforts, and here is where machine learning truly has value.

Analysis:

It will start with the main segment and look into each subsection to understand what impact it has on the objective segment. We will also carry out preprocessing at the required step and incorporate designing tasks. The goal of conducting thorough exploratory analysis is to prepare and clean data for improved machine learning demonstrations to achieve elite performance and summarized models. Therefore, it should start with decomposing and preparing the dataset for expectation.

Modules :

1.Dataset Collection

The information about the employee rating with different types of attributes with empoloyee performance data are collected from the software company.The feedback of client,project managers are taken to the dataset.

2.Data Cleaning

The large dataset contains more noisy and improper data which have to be preprocessed to produce the quality dataset for further pruning.The data is cleaned and processed with initial stage of removing the null values.

3.Exploratory Data Analysis

Exploratory analysis is a process to explore and understand the data and data relationship in a complete depth so that it makes feature engineering and machine learning modeling steps smooth and streamlined for prediction.EDA involves Univarate,Bivariate or Multivariate analysis. EDA helps to prove our assumptions true or false . In other words,it helps to perform hypothesis testing.

4.Machine Learning Modelling

Machine Learning modeling helps to find the best algorithm with the best hyper parameters to achieve maximum accuracy. The dataset is split into 2 variants.70% of records are taken as training data and used to train the machine learning algorithm.The remaining 30% of dataset is applied to testing which helps to predict the process

5.Report

The Data is visualized based on the output of the machine learning algorithm and the data is mapeed with different types of graphs to analyze and visualize the exact data to the user for the prediction of employee performance.Matplot libraries are implemented to map the results based on the user requirements.

METHODS

There are two unique meanings for classification in machine learning. To ascertain whether classes or groups exist in the data, we can gather a set of observations.Alternately, we can be sure that there is a set of classes and that the intent is to develop a rule or collection of rules that will allow us to define a new data point into one of the classes that have already been formed. The previous category is known as Unsupervised Learning and is hence referred to be Supervised Learning[19].

This addressed the classification of the data as supervised learning since it is divided into running and finished sections. The idea of several comparative categorization methods is covered in this section.

Evaluation of the neighboring data points and evaluation of the class system on the adjacent classes are the two stages of classification using KNN. You can calculate the neighbors' distances using methods like Euclidean distance. Class may be determined by a community's majority vote or by analyzing in relation to class in the other way.

The data was scaled to a range of [0, 1] before building the KNN-based model.

VI. DATA DESCRIPTION

Data Description

Variable	Definition
employee_id	Unique ID for employee
department	Department of employee
region	Region of employment (unordered)
education	Education Level
gender	Gender of Employee
recruitment_channel	Channel of recruitment for employee
no_of_trainings	no of other trainings completed in previous year on soft skills, technical skills etc.
age	Age of Employee
previous_year_rating	Employee Rating for the previous year
length_of_service	Length of service in years
KPIs_met >80%	if Percent of KPIs(Key performance Indicators) >80% then 1 else 0
awards_won?	if awards won during previous year then 1 else 0

The Data Description shows all the attributes of the employee dataset of both training and test data

Descriptive Statistics:

1)Descriptive Statistics for the Numerical Columns we check for stats such as Max,Min,Mean,Count,standard deviation,25 percentile,50percentile and 75 percentile.

2)Then Check for the Descriptive Statastics for Categorical Columns for Categorical Columns we check for stats such as count,frequency,top,and unique elements.

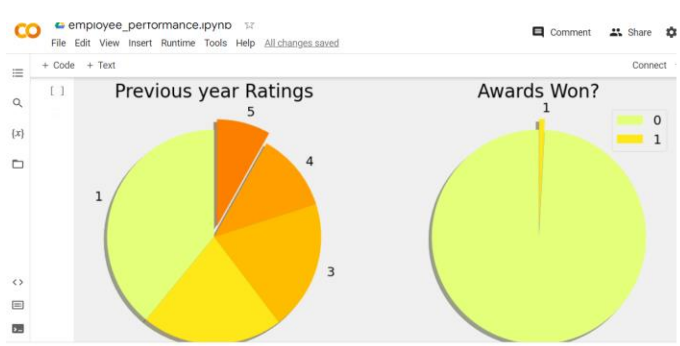
Once the dataset has been put into the Python CoLab, the noisy data are first removed. The dataset needs to be cleaned up because it can be fragmented and can't be provided directly to the model. Therefore, I'll create a cleaning feature that uses the supporting system to clean the data and then outputs the cleaned text:

a) Remove all numbers, Alphanumeric words, such as words that contain the first and second letter of the same word, and numerals, such as hello123.

VII. RESULTS/EXPERIMENTATION & COMPARISON ANALYSIS

Data Visualization :

The employee details with their ratings



The previous year awards and ratings won were takwn into consideration

Result of splitting data:

```
Shape of the x : (3489, 9)
Shape of the y : (3489,)
Shape of the x Test : (23490, 9)
```

The x has 3489 records with 9 columns and y have 3489 records.

This sample's population represents the workforce in the United States as a whole, which is made up of individuals at various phases of their careers and with a variety of educational backgrounds.

rates of pay and performance, as well as people from different backgrounds. Therefore,taking into consideration the numerous themes and classes that naturally exist in the data, it is reasonable to deduce that the most likely result is a rule-based methodology or a tree-based model. This support the hypothesis. It can be seen that the two Random Forest and XGBoost tree-based classifiers outperform the other classifiers during training, and that XGBoost outperforms Random Forest significantly during testing. The accuracy and storage efficiency of the XGBoost classifier are superior to those of the other classification model.

Resampling is the method that consists of drawing repeated samples from the or iginal data samples.The method of Resampling is a nonparametric method of st atistical inference.

The prediction variable are selected for the purpose of the model evaluation with the dataset.

▸ Resampling

```
# It is very important to resample the data, as the Target class is Highly imbalanced.
# Here We are going to use Over Sampling Technique to resample the data.

from imblearn.over_sampling import SMOTE

x_resample, y_resample  = SMOTE().fit_resample(x, y.values.ravel())
```

Results of Re-sampling

```
# lets print the shape of x and y after resampling it
print(x_resample.shape)
print(y_resample.shape)
```

```
(6510, 9)
(6510,)
```

Machine Learning Predective Modelling:

The Decision Tree classifier is applied for the employee performance rating system:

Decision Tree Classifier

```
# Lets use Decision Trees to classify the data
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier()
model.fit(x_train, y_train)

y_pred = model.predict(x_valid)
```

The algorithm is executed and accuracy is calculated

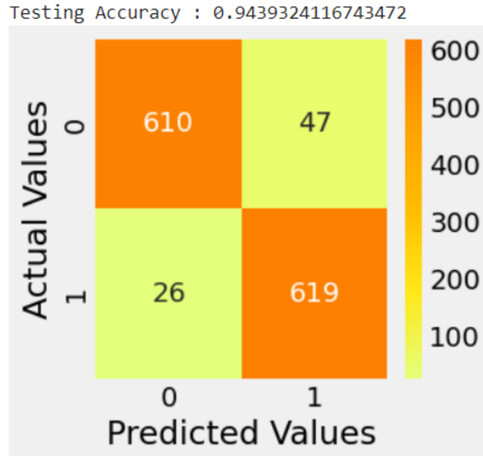
```
from sklearn.metrics import confusion_matrix, classification_report

print("Training Accuracy :", model.score(x_train, y_train))
print("Testing Accuracy :", model.score(x_valid, y_valid))

cm = confusion_matrix(y_valid, y_pred)
plt.rcParams['figure.figsize'] = (3, 3)
sns.heatmap(cm, annot = True, cmap = 'Wistia', fmt = '.8g')
plt.xlabel('Predicted Values')
plt.ylabel('Actual Values')
plt.show()
```

Result:

```
Training Accuracy : 0.9976958525345622
Testing Accuracy : 0.9439324116743472
```



Classification Report:

```
# lets take a look at the Classification Report

cr = classification_report(y_valid, y_pred)
print(cr)
```

```
              precision    recall  f1-score   support

           0       0.96      0.91      0.93       657
           1       0.91      0.96      0.93       645

    accuracy                           0.93      1302
   macro avg       0.93      0.93      0.93      1302
weighted avg       0.94      0.93      0.93      1302
```

The classification report displays the accuracy,precision,recall,f1-score,support.

REFERENCES

[1] Kevser Şahinbaş, " Employee Promotion Prediction by using Machine Learning Algorithms for Imbalanced Datase ", *n September 14,2022 at from IEEE Xplore©2022 IEEE*.

[2]] Jung-gu Choi1*, Inhwan Ko2*, Jeongjae Kim1, Yeseul Jeon3, and Sanghoon Han1,2, " Machine Learning Framework for Multi-level Classification of Company Revenue", *IEEE* 2021.

[3] P. Choudhury, C. Foroughi and B. Larson, "Work- from- anywhere: The productivity effects of geographic flexibility", *Strategic Management Journal*, vol. 42, no. 4, pp. 655-683, 2021..

[4] Richard Josep, Shreyas Udupa and Sanket Jangale, "Employee Attrition Using Machine Learning And Depression Analysis ", *978-0-7381-1327-2/21/$31.00 ©2021 IEEE*.

[5] B. Becker, B. Gerhart, "The Impact of Human Resource Management on Organizational Performance: Progress and Prospects," Academy of Management Journal, vol. 39, no. 4, pp.779–801, Aug. 1996

[6] D. E. Guest, "Human resource management and performance: a review and research agenda," International Journal of Human Resource Management, vol. 8, no. 3, pp. 263–276, Jun. 199

[7] J. Pfeffer, "Seven Practices of Successful Organizations," California Management Review, vol. 40, no. 2, pp. 96–124, Winter, 1998.

[8] J. E. Delery, D. H. Doty, "Modes of Theorizing in Strategic Human Resource Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions," Academy of Management Journal, vol. 39, no. 4, pp. 802–835, Aug. 1996

[9] J. B. Arthur, "Effects of Human Resource Systems on Manufacturing Performance and Turnover," Academy of Management Journal, vol. 37, no. 3, pp. 670–687, Jun. 1994

[10] M. A. Youndt, S. A. Snell, J. W. Dean, Jr., D. P. Lepak, "Human Resource Management, Manufacturing Strategy, and Firm Performance," Academy of Management Journal, vol. 39, no. 4, pp. 836–866, Aug. 1996.

[11] J. P. MacDuffie, "Human Resource Bundles and Manufacturing Performance: Organizational Logic and Flexible Production Systems in the World Auto Industry," Industrial and Labor Relations Review, vol. 48, no. 2, pp. 197–221, Jan. 1995.

[12] M. A. Huselid, "The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance," Academy of Management Journal, vol. 38, no. 3, pp. 635–672, Jun. 1995.

[13] Cristianini, N.; Shawe-Taylor, J. An Introduction to Support Vector Machines and other kernel- based learning methods.Ai Magazine 2000, 22, 190.

[14] C. Cortes and V. Vapnik, Support-vector networks. Machine learning, 20(3), 273-297,1995.

[15] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of computer and system sciences, 55(1), 119-139,1997.

[16] S. Lessmann and S. Voß, "A reference model for customercentric data mining with support vector machines", EuropeanJournalofOperationalResearch199,520–530,2009.T. Fawcett, "An introduction to ROC analysis", Pattern Recognition Letters 27 (8), 861–874,2006.

[17] A. Liaw and M. Wiener, "Classification and regression by randomForest", R news, 2(3), 18-22,2002.

[18] H. A. Elsalamony (2014), "Bank direct marketing analysis of data mining techniques", International Journal of Computer Applications,85(7).

[19] M. Stoval and N. Bontis, "Voluntary performance: Knowledge management– Friend or foe?", Journal of Intellectual Capital, 3(3), 303-322,2002.

[20] S. Jahan, "Human Resources Information System (HRIS): A Theoretical Perspective", Journal of Human Resource and Sustainability Studies, Vol.2 No.2, Article ID:46129,2014.