Employee Classification for Personalized Professional Training Using Machine Learining Techniques and Smote

Mr.I.Mohamed Alfasith[#1], MS.A.Ananthi., M.E., [#2]

[#1]*PG Student,*[#2]*Assistant Professor & Sir Issac Newton College of Engineering and Technology*
Department of Computer Science & Sir Issac Newton College of Engineering and Technology– Nagapattinam
Tamilnadu, India

ABSTRACT:
Training and development are essential parts of professional development that are necessary for employees to improve their capacity. Generally, professional development program is organized based on personal information such as background, personal goal, and work experience together with business objectives and job criterion. To promote personalized training in professional development process, the proper classification of individual employee is necessary. This paper thus proposes the classification method for employee classification to promote personalized training in organizations. The machine learning based method, Decision tree, Random Forest and Support Vector machine, are studied. Synthetic minority Oversampling Technique (SMOTE) method is used to deal with imbalance data. The open data form kaggle is used in this paper. For method validation, the data for training and testing are formed into three gropes including 80:20, 70:30 and 60:40 respectively. The classification results show that the SMOTE can improve classification performance for all classifiers. Additionally, random forest performs the best classification accuracy.
KEYWORDS: Decision Tree, SMOTE, Random Forest and Support Vector Machine, Human Resource Management

I. INTRODUCTION

Human resource management is a method to manage human resource in organization. One main process of human resource management is training and development. Training is a short-term process; it provides learning basic skills and knowledge for each job. On the contrary, development is a long-time process focusing on individual or personal development to improve their performance for future jobs' necessities including conceptual or theoretical inputs, perspective strategic thinking, and behavioural aspects, such as leadership skills, team management, etc. Development process needs employees to handle complex or challenging task. Thus, it spends long-term process.

Professional Development (PD) is a process of development. It uses background knowledge, skills and experience to improve individual capacity through leaning and applying processes. There are many ways of learning and developing specific skills, techniques, knowledge, and abilities. Nowadays, professional development process in many organization pays the important role, since job recruitment process has finished to satisfy business objective or job position criterion.

The process of training and development has many factors which can be occurred form employee problem or organization problem. Especially for organization problems, training and development process normally has high cost, requires time consumption, and challenges with selecting inappropriate employees. Form these evidences, an organization normally select employees who have high professional development process. Therefore, skills or background knowledge of employees need to be evaluated and analysed with job criterions. However, all employees require different training and development process. The appropriate method for selecting right individual to the right training and development process regarding the personal goal and business objective is indeed required, which is the aim of this paper.

Generally, each organization has different techniques to manage human resources professional development. Currently, many organizations use different human resource tools to manage their human resources, depending on the business' objectives to assess employees'

---

performance. The decision-making process depends on the report by analysing relevant attributes such as performance, KPL, length of service, merit and ability, education, technical, potential, training, etc.

AIM

This paper aims to promote the professional training which is the important part of the professional development process. More specifically, this paper proposes the employee classification method for personalized training in the organization. The machine learning based method is studied. At the same time, Synthetic Minority Oversampling Technique (SMOTE), which is one of the methods oversampling a dataset used in a typical classification problem, is used to deal with the imbalance data in this paper. This paper demonstrates the comparison of three machine learning algorithms including Decision Tree, Random Forest, and Support Vector Machine. For the experiment, the open human resources dataset form Kaggle is employed in this paper. The main objective method for employee classification.

II. METHODOLOGY

2.1 Data Collection
The dataset consists of employee records from various departments, including demographics, job roles, performance scores, previous training history, and feedback. The data is pre-processed to handle missing values and normalized for better performance of machine learning models.

2.2 Classification Models
Various machine learning algorithms are employed, including:
- **Decision Trees**: For their interpretability and ease of use.
- **Support Vector Machines (SVM)**: For handling non-linear relationships in the data.
- **Random Forests**: To improve generalization by reducing overfitting.
- **Neural Networks**: For capturing complex patterns in the data.

2.3 Addressing Data Imbalance with SMOTE
SMOTE is applied to the training data to oversample the minority classes. This technique generates synthetic samples by interpolating between existing minority class samples, thus balancing the dataset and improving the model's ability to correctly classify all employee categories.

III. APPLICATIONS AND IMPLEMENTATION

The "Employee Classification for Personalized Professional Training using Machine Learning Techniques and SMOTE" project can be utilized in various ways to enhance professional development and optimize organizational efficiency". Here's how the project can be applied in practice:

3.1 Human Resources Management
- Training Needs Assessment: HR departments can use the classification model to assess the training needs of employees, ensuring that each employee receives the appropriate training to enhance their skills and performance.
- Performance Improvement Plans: Identify employees who may benefit from targeted training programs to improve their job performance.

3.2 Personalized Training Programs
- CustomTraining Pathways: Develop personalized training pathways for employees based on their classification. For example, new hires may receive onboarding training, while seasoned employees might get advanced skill development.
- Skill Gap Analysis: Identify and address specific skill gaps within the organization, ensuring that employees are equipped with the necessary competencies.

3.3 Career Development
- Career Path Planning: Assist employees in planning their career paths by recommending relevant training and development opportunities that align with their career goals.
- Mentorship Programs: Pair employees with mentors based on their classification and training needs to foster professional growth.

3.4 Data-Driven Decision Making
- Resource Allocation: Allocate training resources more effectively by understanding the distribution of training needs across the organization.
- Strategic Planning: Use insights from the classification model to inform strategic decisions regarding talent development and workforce planning.

3.5 Monitoring and Evaluation
- Training Effectiveness: Continuously monitor the effectiveness of training programs by evaluating the performance of employees post-training. Adjust training content and delivery methods based on feedback and performance data.
- Employee Progress Tracking: Track the progress of employees over time, ensuring that they are meeting their professional development goals.

3.6 Enhanced Employee Engagement
- Personalized Feedback: Provide personalized feedback and development recommendations to employees, increasing their engagement and satisfaction.
- Recognition and Rewards: Recognize and reward employees who successfully complete training programs and demonstrate improved performance.

3.7 Balancing Training Opportunities
- Equitable Training Access: Ensure that all employees, including those in minority classes, have equal access to training opportunities by using SMOTE to balance the training dataset.

3.8 Deployment and Integration
- Integration with HR Systems: Integrate the classification model with existing HR systems and Learning Management Systems (LMS) to streamline the process of training needs assessment and program deployment.
- User-Friendly Interface: Develop a user-friendly interface for HR professionals and managers to easily access and utilize the classification and recommendation system.

IV. PROPOSED SYSTEM

In our predicted model, ten features have been evaluated to make this comparison more unique. Our introduced algorithms were conducted based on SMOTE the obtained outcomes were compared to other works to show the percentage of improvement, while decrease in performance also noted in one occasion Random Forest. The highest increment was noticed for previous works which was about percentage improvement were calculated for Employee promotion is used to determine whether an Employee promoted are not promoted to know this we used the machine learning based methods such as Decision Tree, Random Forest and Support Vector Machine classification techniques to figure out.

4.1 Advantages
- High efficiency
- Time Saving
- Inexpensive
- Low complexities
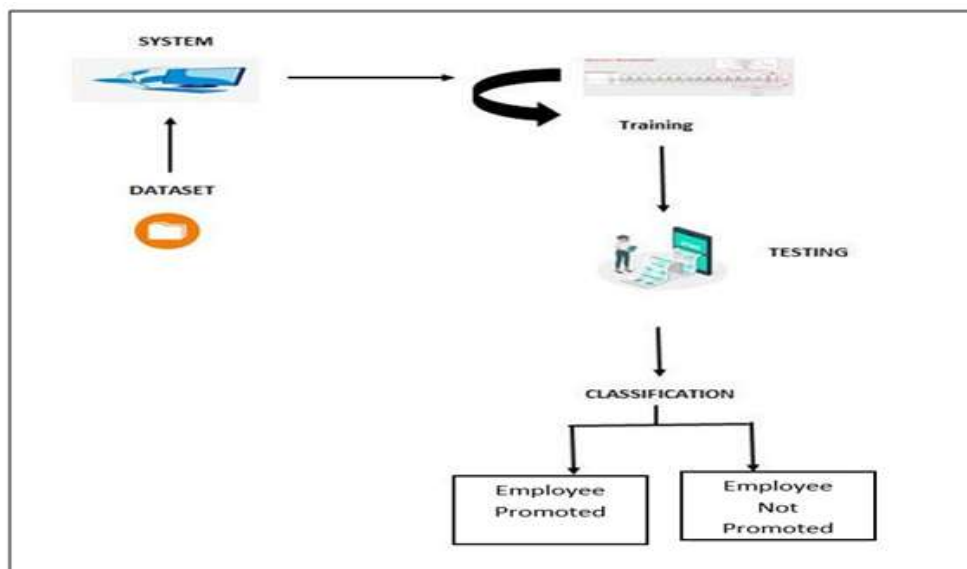
4.2 Proposed Architecture

Fig 3.2 SYSTEM ARCHITECTURE

Fig 3.2 represents a process flow diagram for an employee promotion prediction system using machine learning techniques. The system aims to classify employees into two categories: "Employee Promoted" and "Employee Not Promoted.
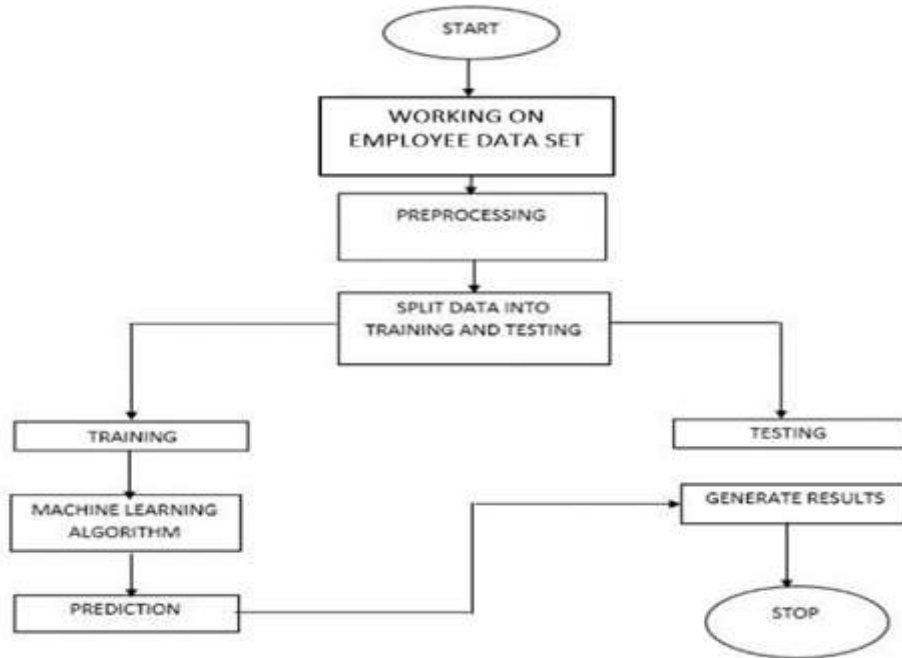
4.3 Block Diagram

Fig 3.2.1 BLOCK DIAGRAM

The Fig 3.2.1 presents a flowchart depicting the steps involved in predicting employee promotions using a machine learning approach. The process begins with data handling and ends with generating prediction results.

V. MODULES DESCRIPTION
1. System
2. User

1. System
1.1 Receive Datasets:
Receive Datasets from the user

1.2. Pre-processing:
Perform pre-processing on data sets

1.3 Training:
Use the pre-processed training dataset to train our models

1.4 Generate Results:
View generated Results

2. User
2.1 Register:
Users can register for the Employee Classification for Personalized Professional Training Using Machine Learning Techniques and Smote application here.

2.2 Login:
After registering, the user can access his portal.

2.3 Upload:
The user needs to upload the data.

2.4 View-Data:
Later on user can upload data

2.5 View Pre-processing:
Later on user can view the pre-processing of data.

2.6 View training:
Later on, user can view the training of data.

2.7 View Prediction:
User need to input the required fields in order to get the response from the system.

VI. RESULT ANALYSIS

The models demonstrate a high accuracy in predicting the majority class (e.g., high performers or well-trained employees), but significantly lower precision and recall for minority classes (e.g., low performers or those with specific training needs). The following key observations were made:

- **Decision Trees and SVMs**: High accuracy but struggled to identify underrepresented classes, leading to skewed results.
- **Random Forests**: Better performance in handling data variance but still biased toward the majority class.
- **Neural Networks**: Showed potential in capturing complex patterns but overfitting on the majority class was observed.

Overall, the results indicate a need for techniques like SMOTE to balance the dataset and improve classification accuracy for all employee categories.The use of machine learning and SMOTE proves to be effective in classifying employees for personalized training. The improved performance post-SMOTE application highlights the importance of addressing class imbalance in employee data. This approach can be integrated into human resource management systems to automate the classification and training process, thereby enhancing workforce development.

VII. CONCLUSION

In this research we have developed a web application where it results an employee is eligible for promotion or not. We have calculated this based on few factors i.e., Experience, Previous year ratings, KPI's performed, Average score of trainings and No. of trainings attended by the employee. This web application also considered basic factors like gender, age, region, education and department of an employee to calculate and show whether employee is eligible for promotion or not. To perform this we have used machine learning models of Decision Tree, Random Forest and Support Vector Machine (SVM).

REFERENCES

[1]. Y. Asim, B. Raza, A. K. Malik, S. Rathore and A. Bilal," Improving the Performance of Professional Blogger's Classification", 2018. International Conference on Computing, Mathematics and Engineering Technologies (iCOMET), Sukkur, 2018, pp. 1-6.

[2]. T.W Ramdhani, B. Purwandari and Y. Ruldeviyani, "The Use of Data Mining Classification Technique to Fill in Structural Positions in Bogor Local Government" 2016 International conference on Advanced computer Science and Information Systems (ICACSIS), Malang 2016.

[3]. V. Mathew, A.M. Chacko and A. Udhayakumar, "Prediction of suitable human resource for replacement in skilled job positions using Supervised Machine Learning "2018 8th International Symposium on Embedded Computing and System Design (ISED), Cochin, India, 2018.

[4]. D. Hartanto, E. Utami, S. Adi and H.S.Hudnanto," Job Seeker Profile Classification of Twitter Data Using the Naïve Bayes Classifier Algorithm Based on the DISC Method", 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering(ICITISEE), Yogyakarta, Indonesia, 2019.

[5]. T. Tarusov and O. mitrofanova, "Risk Assessment in Human Resource Management Using Predictive Staff Turnover Analysis", 2019 1st International Conference on Control System, Mathematical Modelling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russia, 2019.

[6]. Qiangwei Wang, Boyang Li and Jinglu Hu, "Feature Selection for Human Resource Selection Based on Affinity Propagation and SVM Sensitivity Analysis," 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, 2009.

[7]. M. Eminagaoglu and S. Eren, "Implementation and comparison of machine learning classifiers for information security risk analysis of a human resources department,' 2010 International Conference on Computer Information System and Industrial Management Applications (CISIM), Krakow, 20010.

[8]. L.I.FDutsinm and P. Temdee,"VARK Learning Style Classification Using Decision Tree with physiological signals," 2020 Wireless personal Communication, 2020.

[9]. Q.Guohao, W. Bin and Z. Baoil,"Competency Analysis in Human

Resources Using Text Classification Based on Deep Neural Network," 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC).

[10]. N.Aottiwerch and U. Kokaew, "The analysis of matching learners in pair pro gramming using K-means," 2018 5th International Conference on Industrial and Applications (ICIEA), Singapore, 2018.