IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

A REVIEW PAPER: TP AND TPS USING MACHINE LEARNING

[1]Leena Suresh More, [2]Dr. Binod Kumar
[1]PhD Scholar, [2]Professor, PhD Guide
[1]SVIMS, Koregaon Park, Pune
[1]Savitribai Phule Pune University, Pune, India

Abstract: Artificial intelligence (AI) and deep learning have become indispensable tools in various sectors of human resource management, enabling advanced data analytics and prediction capabilities. These technologies facilitate the identification and extraction of novel patterns from vast datasets, a task often impractical using traditional methods. By leveraging AI and deep learning, organizations can streamline the process of talent acquisition by accurately predicting optimal candidates from extensive document pools. Moreover, they can effectively mitigate attrition risks by precisely segmenting documents to identify and promote top performers. This paper highlights the transformative potential of AI and deep learning in enhancing HR management practices.

***IndexTerms* - Machine Learning, Top Performer Segmentation, Talent Prediction.**

INTRODUCTION

Artificial intelligence (AI) has permeated every sector, serving as an expert assistant to simplify human tasks. The human resource department is no exception, as AI offers solutions to harness high-quality HR data effectively. Many industries possess rich HR data but struggle to leverage it optimally to mitigate risks, maximize returns, and forecast workforce trends. The costs associated with attrition, suboptimal hiring decisions, and compensation management are substantial, underlining the need for data-driven insights over traditional judgment-based HR practices.

The ability to predict the potential success of new hires based on their profiles presents a significant challenge for organizations. High attrition rates can further exacerbate this challenge, particularly for rapidly growing companies. Given the sheer volume of candidate profiles, organizations often lack the time to manually review each one. Hence, there arises a critical need for advanced machine learning algorithms to predict talent and forecast training needs accurately. Additionally, segmenting top performers based on employee profiles and performance indices is crucial for effective promotional activities.

Key questions arise regarding the efficacy of existing classifiers in making reliable predictions and the necessity of refining or designing new talent acquisition and management models. Enhancing existing classifiers for improved prediction and segmentation, as well as considering additional factors to enhance model efficiency, becomes imperative. This introduction sets the stage for exploring the potential enhancements and innovations in talent prediction and management using AI and machine learning algorithms within the HR domain.

REVIEW OF LITERATURE

Xiaohua Zeng et al. (2022) investigated the improvement in forecasting stock index using a hybrid model incorporating variables such as WT, LSTM, AGA, and DJIA. Their study revealed an overall enhancement in forecasting accuracy with the AGA-LSTM model, although it fell short of expectations.

Jisoo Ock (2022) found evidence suggesting that resume screening decisions may exhibit bias against applicants from demographic minority groups. Despite adjustments in selection criteria, less weight was given to prediction tests after the initial screening process.

Erzsebet Frigo et al. (2022) proposed an exponential weighted algorithm based on a grid over the space of combination weights, which achieved close to optimal empirical performance for two base rankers while scaling well with an increased number of models. Darcy A. B. Jones et al. (2021) introduced a novel tool and pipeline for ranking predicted effector candidates, interfacing with multiple software tools and methods.

U. Ajaykumar, K. Devasenapathy (2021) emphasized the necessity for better prediction algorithms and explainability in educational data mining, as the existing prediction methods were deemed insufficient.

Arushi Gupta et al. (2021) conducted a review to understand the pipeline and shortcomings of news classification processes, ultimately recommending SVM models for their high accuracy and low training time.

G. Angeline Prasanna, P. J. Anu (2021) proposed data extraction tools to predict learning disabilities in school-age children, highlighting decision trees as a powerful tool for classification.

M. Karmakar et al. (2021) developed an applicant personality prediction system based on resume and test data, utilizing simple keyword-based techniques for filtering.

Ghazal Rafiei et al. (2021) proposed a recruitment process using machine learning to recommend the best candidates according to job descriptions.

Lami Mostafa et al. (2021) devised a job candidate ranking model utilizing SVM, HMM, and NLP approaches, although limitations were identified in handling large datasets and homonyms/acronyms.

Huichao Xue et al. (2020) implemented a job recommendation system using tf-idf and logistic regression to improve precision and recall, but encountered limitations in achieving desired accuracy levels.

Shreya Sawleshwarkar et al. (2018) developed an automated process incorporating psychometric tests, using text mining and scoring mechanisms for candidate shortlisting, but accuracy suffered due to the use of Boolean logic.

Vinay Dandwani et al. (2017) employed a resume building and ranking system based on keyword-based search, efficient for extracting candidates from social network sites but lacking handling for active candidates.

Qinbao Song et al. (2013) utilized fast clustering-based feature subset selection, leveraging graph-theoretic clustering methods to reduce dimensionality, though limited to microarray data.

Kexin Zhu and Jian Yang (2013) adopted an affinity propagation-sequential feature selection approach, finding it faster than traditional sequential feature selection, but with limitations in achieving desired accuracy levels.

PROBLEM STATEMENT:

Organizations across various industries routinely face the challenge of sifting through thousands of resumes for each job posting, necessitating dedicated officers for resume screening. Selecting the right talent amidst this influx is a daunting task, especially for rapidly growing businesses experiencing high attrition rates. Human resource departments often find themselves overwhelmed in such dynamic markets.

In service-based organizations, professionals with diverse domain expertise and technical skills are recruited and assigned to specific projects. The process of candidate screening involves identifying the most suitable resumes or talents from a pool of applicants.

Furthermore, during performance appraisal cycles, employees submit appraisal forms, presenting another challenge for HR teams to manually review each document and identify top performers deserving of promotions or salary hikes.

Previous research on ranking and forecasting applicants' performance has yielded inconclusive results, indicating a need for enhancements to existing models. In light of these challenges, there is a pressing need to explore personalized and efficient approaches or models for predicting desired applicants based on personality assessments, forecasting talent for optimal training, and segmenting top performers for promotion and salary adjustments.

Therefore, the research question at hand can be framed as follows:

"Is there a personalized and efficient approach or model to predict the most desirable applicants, incorporating personality assessment and talent forecasting for optimal training, while also segmenting top performers based on performance indices for promotion and salary hikes?"

ANALYSIS AND FINDINGS:

Previous research has explored various techniques such as SVM, HMM, NLP, regression, and keyword-based search for ranking purposes. However, each model has its limitations, including lengthy training times, lower accuracy rates, inability to handle homonyms and acronyms, and lack of predictive capabilities for active candidates.

Methods like Graph-Theoretic clustering have shown promise in reducing dimensionality, but they are limited to microarray data and struggle with big data. Similarly, the Conditional Dynamic Mutual Information Future Selection model has demonstrated better performance but is sensitive to noise in the data.

The Evolutionary Local Selection algorithm, which employs K-Means clustering, covers a wide space of feature combinations but suffers from decreased cluster quality with an increase in the number of features. Wrapper-based feature selection using SVM offers improved accuracy and faster computation but requires more time for the training phase.

Hybrid feature selection approaches, such as Mutual Information with Model-Based feature selection algorithms, enhance accuracy but come with a high computation cost for high-dimensional datasets.

Overall, while these methods have shown promise in addressing various aspects of the talent prediction and ranking process, each has its trade-offs and limitations. Finding a personalized and efficient approach that balances accuracy, computational efficiency, and scalability remains a key challenge in HR analytics.

Table 1: Comparative analysis of various algorithms

Algorithm Used	**Approach/Method used**	**Outcomes**	**Limitations**
Job candidate rank model	SVM, HMM, NLP	Better accuracy	Problem with large dataset, can't handle homonyms and acronyms
Job recommendation and ranking system	Logistic regression	Improves precision and recall, feature reduction	Accuracy not up to the mark
Resume building and ranking system	Keyword-based search	Efficient to extract candidates from social network sites	Active candidates not handled
Fast clustering based feature subset selection	Graph-Theoretic clustering method	Dimensionality is reduced	Well work with only micro array data
Condition dynamic mutual information future selection	Mutual information	Better Performance	Sensitives to noise
Affinity Propagation-Sequential feature Selection	Affinity Propagation Clustering algorithm with SFS	Faster than sequential feature selection	Accuracy is not better
Evolutionary local selection algorithm	K- Means Algorithm	Covers large space of possible feature combinations	As no of features increase cluster quality decreases
Wrapper based feature selection using SVM	Sequential selection forward with SVM	Better accuracy and faster computation	Takes more time for training phase
Two phase features	ANN weight	Handles both	Takes more time for
selection approach	analysis used to remove irrelevant features with Genetic algorithm to remove redundant features	irrelevant and redundant features. Improves accuracy	training phase and Requires more memory
Hybrid feature selection	Mutual information with model based feature selection algorithm	Improves accuracy	High computation cost for high dimensional datasets

Various metrics can indeed be used to evaluate the performance of a classifier, each shedding light on different aspects of its effectiveness. Here's a brief overview of some commonly used metrics:

1. Accuracy: This metric measures the overall correctness of the classifier's predictions. It's calculated as the number of correct predictions divided by the total number of predictions made. While accuracy provides a straightforward measure of performance, it may not be the best indicator in scenarios where the class distribution is imbalanced.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

2. Precision: Precision focuses on the proportion of correctly predicted positive instances (true positives) among all instances predicted as positive (true positives and false positives). High precision indicates that there are fewer false positives.

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

3. Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances (true positives) among all actual positive instances (true positives and false negatives). High recall indicates that there are fewer false negatives.

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

4. F1 Score: The F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is particularly useful when there is an imbalance between the classes. A higher F1 score indicates better overall performance of the classifier.

$$\text{F1 Score} = \frac{2\times P\times R}{P+R}$$

where P is precision and R is recall.

These metrics collectively offer insights into different aspects of the classifier's performance, allowing for a comprehensive evaluation of its effectiveness in making predictions across various categories.

TABLE 2: CLASSIFIERS PERFORMANCE EVALUATION BASED ON VARIOUS METRICS

Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic regression	94.73%	0.86	0.75	0.80
KNN	99.42%	0.94	0.94	0.94
SGD	90.22%	0.47	0.23	0.31

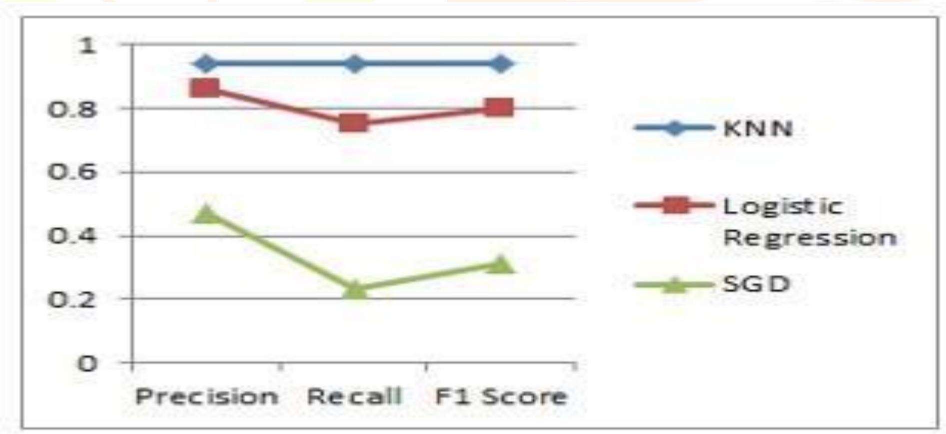

FIGURE 1: GRAPHICAL REPRESENTATION OF VARIOUS ALGORITHMS EVALUATION PERFORMANCE BASED ON PRECISION, RECALL AND F1 SCORE

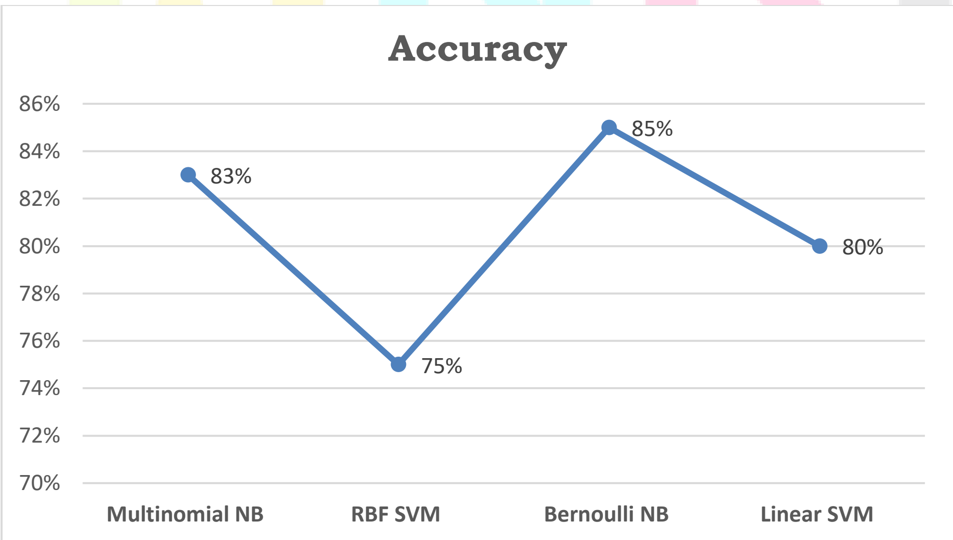

Figure 2: Graphical representation of various algorithms evaluation performance based on accuracy

TABLE 3: CLASSIFIERS PERFORMANCE EVALUATION BASED ON VARIOUS METRICS

Algorithm	Accuracy	Precision	Recall	F1 Score
Multinomial NB	83%	0.47	0.71	0.57
RBF SVM	75%	0.41	0.31	0.35
Linear SVM	80%	0.51	0.61	0.56
Bernoulli NB	85%	0.57	0.71	0.63
Logistic regression	79%	0.5	0.6	0.55

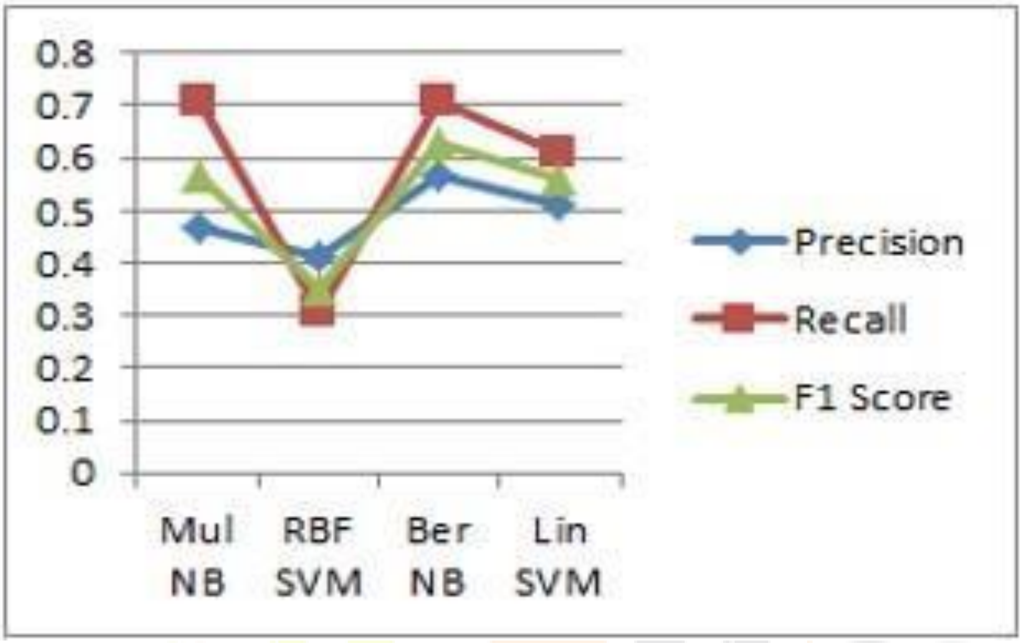

Figure 3: Graphical representation of various algorithms evaluation performance based on precision, recall and F1 Score

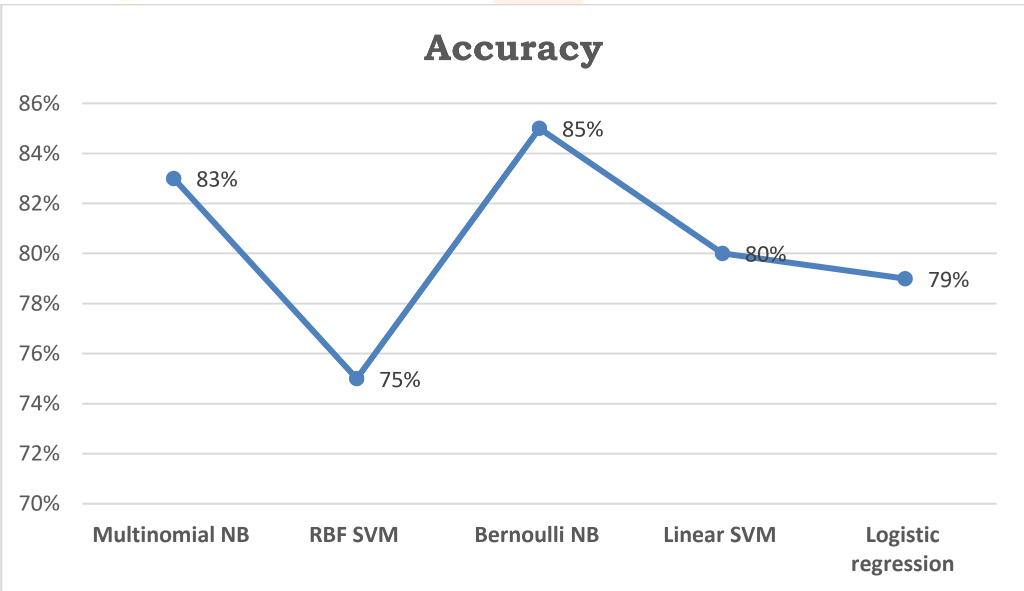

FIGURE 4: GRAPHICAL REPRESENTATION OF VARIOUS ALGORITHMS EVALUATION PERFORMANCE BASED ON ACCURACY

INTERPRETATION AND CONCLUSION:

Based on the analysis presented in Table 2 and Graphs 1, 2, it is evident that the Bernoulli Naive Bayes classifier exhibits higher accuracy, precision, recall, and F1 score compared to other classifiers when considering the given document tags. Conversely, Table 3 and Graphs 3, 4 indicate that the K-Nearest Neighbors (KNN) classifier outperforms others in terms of accuracy, precision, recall, and F1 score based on the same document tags.

In comparison to existing classifiers, the proposed ensemble classifier offers several advantages. It effectively handles both irrelevant and redundant features extracted from various documents uploaded by applicants. This capability contributes to improved prediction accuracy and classifier performance while enabling faster computation. The proposed classifier demonstrates efficiency in accurately predicting talent according to the requirements of the job profile. Additionally, it enhances personality prediction capabilities compared to existing methods and facilitates better talent forecasting for training purposes.

Moreover, the proposed classifier assists in segmenting top performers based on existing employee performance appraisals. Unlike existing models that rely on boolean logic, the proposed classifier operates on fuzzy logic, allowing for nuanced weighting of parameters ranging from 0 to 100%. This mimics the decision-making process of human resource managers more closely, thereby enhancing the overall efficacy of talent prediction and management processes.

In conclusion, the proposed ensemble classifier offers a promising solution to address the challenges associated with talent prediction, personality assessment, training forecasting, and top performer segmentation in human resource management. Its superior performance metrics and advanced functionality position it as a valuable tool for organizations seeking to optimize their HR practices and make more informed decisions regarding talent acquisition and management.

REFERENCES

1. Xiaohua Zeng, Jieping Cai et al., "A hybrid model integrating long short-term memory with adaptive genetic algorithm based on individual ranking for stock index prediction", Springer, www.doi.org/10.1371/journal.pone.0272637, 2022.

2. Erzsebet Frigo, Levento Kocsis, "Online convex combination of ranking models", Springer,www.doi.org/10.1007/s11257-021-09306- 7,2022.'www.mdpi.com/journal/sustainability, 2022.

3. Darcy A. B. Jones, Lina Rozano et al., "An automated and combinative method for predictive ranking of candidate effector proteins of fungal plant pathogens", Nature Portfolio, doi.org/10.1038/s41598-021-99363-0, 2021.

4. Lamiaa Mostafa et al., "Job candidate rank approach using machine learning techniques", ACM International conference on information and knowledge management, 2021.

5. Aarushi Gupta et al., "Unravelling modern news classification methods: A Systematic review", Journal of computer technology & applications, 2021.

6. U. Ajaykumar, K. Devasenapathy, "A study on prediction and explainability in educational data mining", KARPAGAM Journal of computer science, 2021.

7. M. Karmakar et al., "Applicant personality prediction system using machine learning", https://ieeexplore.ieee.org/document/9587693, 2021.

8. G. Angeline Prasanna, P. J. Anu, "Introduction of decision tree technique of data mining to predict learning disability in school-age students", KARPAGAM Journal of computer science, 2021.

9. Ghazal Rafiei et al., "Towards automating human resource recruiting process", ieeexplore.ieee.org/document/9690504, 2021.

10. Huichao Xue et al., "Ranking user attributes for fast candidate selection in recommendation system", www.researchgate.net, 2020.

11. Shreya Sawleshwarkar et al., "Simplified recruitment model using text-mining on psychometric and aptitude tests", https://ieeexplore.ieee.org/document/8474769, 2018.

12. Vinay Dandwani et al., "Candidate ranking and evaluation system based on digital footprints", IOSRJCE, 2017.

13. Hwang, Young-Sup, "Wrapper-based feature selection using support vector machine", Department of Computer Science and Engineering, Life Science Journal, 2014.

14. Qinbao Song et al., "A fast clustering-based feature subset selection algorithm for high-dimensional data", IEEE Transactions on Knowledge and Data Engineering, Vol 25, No 1, 2013.

15. Kexin Zhu and Jian Yang, "A cluster-based sequential feature selection algorithm", IEEE, 2013.