International Journal of Research in Human Resource Management

E-ISSN: 2663-3361
P-ISSN: 2663-3213
IJRHRM 2022; 4(2): 31-36
Received: 19-05-2022
Accepted: 08-07-2022

Efstratia Stasi
Hellenic Open University
Patras, Greece

Georgios Rigopoulos
a) Hellenic Open University
Patras, Greece
b) National and Kapodistrian University of Athens, Athens, Greece

Corresponding Author:
Efstratia Stasi
Hellenic Open University
Patras, Greece

Naïve Bayes classifier for optimizing personnel selection process in financial industry

Efstratia Stasi and Georgios Rigopoulos

Abstract
Algorithmic human resources management (HRM) is becoming increasingly popular among organizations and many HRM processes include automated decision making HRM functions. Research is very active in the domain, and spans across machine learning and data mining, aiming to provide accurate methods to predict best candidates for job roles, or for personnel development among others. In this work, we present a Naïve Bayes based model, which focuses on the preliminary application screening steps, and suggest suitable applicants for further processing, based on a number of features. The model is presented, along with an application in a real case worked with a financial organization and using primary data selected from candidate applications. The results are promising and demonstrate that a mix of professional expertise along with algorithmic support may optimize the HMR processes.

Keywords: Naïve Bayes classifier, human resources, financial sector.

Introduction

Nowadays, Human Resources Management (HRM) is witnessing a systemic transformation from a traditional personnel management approach, to a new, more strategic form, of business- partner management. The traditional HRM has given way to a more reactive HRM, that is expected to be more inclusive, agile, and responsive. This transition is enabled by artificial intelligence (AI) and machine learning (ML) developments, which reduce cost and save time in routine processes, while also enable predictive capabilities for complex processes [1]. Technological innovation, and development of algorithmic technologies has allowed, employers and human resources professionals to take real-time and more effective decisions. Automated or algorithmic decision-making systems have increasingly been used by corporations taking advantage of the huge amount of data generated in contemporary working environment. However, even business benefits of automation in HRM are obvious, utilization of algorithmic systems has come under scrutiny. Ethical aspects, non-explainable AI, bias, and lack of transparency in the decision-making process, are some of the concerns raised from researchers and professionals [2]. In the era of big data, however, where large volumes of data, like hundreds of job applications, have to be processed in limited time, utilization of HR decision support systems and algorithms seems to be a necessity, and strategic choice for organizations, which can lead to improved HR operations.

In this context, present work focuses on automated decision support systems in Human Resources Management. Specifically, it presents an algorithmic model for hiring decisions, based on Naïve Bayes algorithm. The model was built and tested using real HR data, originated from the financial industry. The collected data were anonymized, and transformed into the appropriate format, but no further information can be disclosed, due to non-disclosure agreement. The work demonstrates that the application of Naïve Bayes classifier is an efficient approach in developing a decision support system for HR professionals for the employee selection process.

The rest of the paper is organized as follows. Initially, literature review introduces HR algorithmic decisions and the challenges that are encountered, such as ethical and legal considerations, data protection, explain ability, and accountability issues. Next, the Naïve Bayes model is presented, including design and testing, along with results. Finally, discussion follows with main findings and suggestions for further improvements.

Literature review

Research in concepts such as people analytics, human resource management algorithms, and algorithmic control is becoming very popular In Human Resources Management discipline. In addition, these concepts are studied in parallel, and sometimes interchangeably, with developments in big data and artificial intelligence [3]. To synthesize all these concepts Meijerink *et al*. (2021) [3], propose the term “Algorithmic Human Resources Management (HRM)” and define it as: “The use of software algorithms that operate based on digital data to augment HR-related decisions and/or automate HRM activities.”

The term algorithmic HRM integrates insights on issues, such as the increasing usage of data to support HR decision-making, development of software algorithms that process working data, and partial or full automation of HR decision-making, all of which have significant impact on how labor is managed and HRM practices are carried out [3].

Cheng *et al*. (2019) [4] define HRM algorithms as “computer programs of a heuristic nature, that use economical input of variables, information, or analytical resources to approximate a theoretical model, enabling an immediate recommendation of screening, selection, training, retention, and other HR functions”, and separate them into deterministic and probabilistic. Deterministic algorithms assume a deterministic relationship between inputs and outputs, which essentially means that if an input A results in an output B, then A should always be followed by B. While, probabilistic algorithms are used to reveal probabilistic relationships between inputs and outputs, which means that the occurrence of A increases the probability of B.

Such algorithms are already embedded in HR decision-making tools, which differ from the more traditional ones in that they monitor contextual performance, such as employee engagement, integrate data from different sources, and have expanded technical capabilities for analyzing data [5]. It is noteworthy, however, to mention that many practitioners are using the term "black box" for algorithms, particularly when they are concerned that these applications are generating non explainable results [4].

Cheng et.al (2019) [4] research revealed that HR functions that have attracted more interest from human resources management practitioners are recruitment and selection, training and development, and compensation. All, can benefit from algorithmic support as follows:

1. Recruitment is the process where an organization seeks applicants for potential employment [6]. It refers to any action undertaken by a business with the primary goal of identifying and attracting potential employees [7]. Machine learning (ML) can support this activity in evaluating candidate suitability; extracting information from resumes and analyzing applicants' profiles [1].
2. Selection refers to the process where an organization attempts to identify applicants with the necessary knowledge, skills, abilities, and other characteristics that will help the company achieve its goals [6]. Machine learning (ML) in the selection process has led to efforts to identify decision attributes to use as selection criteria and to develop corresponding selection models [1].
3. Training and development consists of all activities that an organization undertakes to facilitate the acquisition of -related knowledge, skills, and behaviors by its employees to meet the challenges [6]. Algorithms can assist in identifying employees’ training needs and recommending relevant courses for career development [1, 8].

Despite the promising and wide utilization of artificial intelligence in human resources, there also exist various challenges. These include [9].

1. Complexity of HR processes.
2. Small data sets.
3. Ethical and legal considerations.
4. Employee reactions to AI.

Also, Garg *et al*. (2021) [1], point out the following challenges arising from the implementation of automated decision-making in all aspects of human resources

1. Availability and sufficiency of data serving as training data for algorithms used for decision making.
2. Data authenticity resulting from the usage of external data, such as social media.
3. Security of employee data stored in company cloud services.
4. Explain ability, fairness, and accountability of algorithms.

Employee selection is among the most important HRM practices that support an organization in its strategy execution. Selection is the process where an organization identifies applicants that possess the necessary skills, expertise, and abilities to help the organization achieve its goals. There are several approaches in recruiting and selecting. Some organizations may actively recruit from external applicants, whereas others rely primarily on internal applicants with the necessary skills to fill vacant positions [6]. Employee selection process differs among organizations and job openings, however, for the majority of organizations, the procedure includes a set of identical steps. In particular, in the first step, a human resources expert screens the applications to determine which of them satisfy the minimum job requirements. In the next steps, the company runs tests and views their work examples, to check if candidates meet the basic criteria and further invite them for interviews. Following the interviews, HR decision-makers begin to short list candidates which are fitting the expected roles. Before the final selection, HR decision-makers check references and conduct background checks, for the top listed candidates [7]. This process, mainly performed by HR professionals for decades, nowadays it has been affected by automation and information technology advances, and algorithmic support.

Proposed Model

Following the above developments, this work considers that utilization of AI can assist HRM operations, given that limitations are taken into account. It proposes a model that can enhance HR decision support decisions in employee selection process. Our model aims to support HR experts in application screening, the first step of the selection process, to determine which of the applicants satisfy the minimum job requirements. The problem of predicting which job applicants will be qualified for further evaluation by human resource experts can be structured as a classification problem. So, the objective is to build a classification model that, based on historical data and applicant features, predicts which applicants -for specific job postings- are likely to pass the first step of screening application. Some well-known classification models in machine learning are Decision Trees (DT), Naïve Bayes (NB), Rule-based method, K-nearest neighbors (KNN), Support Vector Machines (SVM), Neural Networks (NN), Random Forest (RF), and Logistic Regression (LR) [10].

The proposed classifier is based on Naïve Bayes algorithm, as it includes some properties, which make it suitable for HR datasets [10], as follows:

1. Data contain categorical and nominal variables and hence Naïve Bayes could be applied for both types.
2. NB algorithm works well with small training datasets to determine the necessary classification parameters.
3. It simplifies predictive modeling models and it is relatively simple to code.
4. Bayes learning algorithms work well with noisy or missing data and could provide probabilistic predictions when appropriate

However, a few relevant works use NB in HR problems in similar setting, like Khairina *et al*. (2017) [10]. In their study, they used the Naive Bayes algorithm as a method to create a decision support framework for the selection and placement of human resources, according to defined company's criteria, related to education, GPA, interview, age, and experience.

Naïve Bayes Classifier

Naïve Bayes classifier is a popular algorithm that is based on Bayes' probability theorem of conditional independence. Naïve Bayes assumes by design that all attributes are independent of each other, given the class. The objective of a Bayesian classifier is to choose the most likely class from a list of potential ones. Bayes theorem computes the probability of an event B occurring, given the probability of another already occurred event A [12] as follows:

$$(B \setminus A) = \frac{P(A \setminus B)P(B)}{P(A)} \quad (1)$$

where P(B\A) is the probability of B given A, B represents the dependent event, and A the prior event.

The Bayes equation can be written as

$$Posterior = \frac{Likelihood * Prior}{Evidence} \quad (2)$$

Where the factors Likelihood and Prior are functions and Evidence is a constant [12, 13].

Naïve Bayes, computes the conditional probability for each decision class, given an information vector. The probabilities involved in producing the final estimate are computed as frequency counts from the decision table [14]. For a data sample with an unknown class, the Naïves Bayes classifier will predict that belongs to the class with the highest posterior probability [12].

HR Selection Proposed Model

The model we propose, comprises a series of steps, as detailed below, and can be embedded in a HR decision support system as a classifier, for the initial steps of candidate selection process.

1. **Data collection**: In this step, appropriate data should be collected and formatted in usable form, so as to be used for algorithm training and test. In HR environment, this could be historical data of past job openings and corresponding HR evaluations and decisions, or artificial datasets, based on the requirements.
2. **Training**: In this step the NB classifier is trained, using appropriate method. This is an iterative process, until the optimum accuracy level is reached.
3. **Testing**: In this step the NB classifier is tests for accuracy and hyperparameter value setting. This is an iterative process, until the optimum accuracy level is reached.
4. **Evaluation**: We evaluate the classifier, and repeat the previous steps if the level of accuracy is not the desired. For the evaluation of the classification model, the metrics we utilized were Accuracy, Error rate, Precision, Recall, and F1 score, measured on the test data set:

- Accuracy is a common performance metric that measures the overall success rate of the model. More specifically, is the proportion of classifications predicted correctly and is given by the following equation (5) [11].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Error rate

$$Error\,rate = 1 - Accuracy \quad (6)$$

- Precision is the proportion of outcomes predicted to be positive that are actually positive, or else the fraction of retrieved items that are relevant [11].

$$Precision = \frac{Number\ of\ relevant\ items\ retrieved}{Total\ number\ of\ items\ retrieved} = \frac{TP}{TP+FP} \quad (7)$$

- Recall or true positive rate is the proportion of actual positives that are predicted to be positive, or else the fraction of relevant items that are retrieved out of all relevant items [11].

$$Recall = \frac{Number\ of\ relevant\ items\ retrieved}{Total\ number\ of\ relevant\ items} = \frac{TP}{TP+FN} \quad (8)$$

- F1 score is used as a balance between precision and recall, as it is difficult to obtain high values in precision and recall at the same time (if the precision increases then the recall decreases). The F1 score is utilized to obtain a suitable combination of the Precision and Recall (equation 9)

$$F1 = \frac{2Precision \times Recall}{Precision+Recall} \quad (9)$$

5. **Utilization**: After the successful training and testing steps, the classifier can be used in real data, and it can become part of the HR processes.

The above process we propose as a model, indicates the fact that there is no universal solution that fits to all HR jobs, or roles, and it is necessary to perform the training to adapt to each setting. This might include industry, roles, or other specificities. However, the model is highly adaptable and can be used in a diversity of cases. In the next section, we demonstrate its application in a real case selection process, and present the results, using real dataset, provided from a financial institution and formulated accordingly.

Model Application in HR Selection Process

The Naïve Bayes classifier method, introduced in the previous section, was applied in financial HR setting, and trained on a dataset with records of internal applicants for specific roles, and HR decisions on whether applicants were shortlisted or not. The aim was to train the model, tune the parameters, in order to achieve the highest accuracy level, and embed it in the existing HR processes, as an assistive tool for application screening. The dataset was extracted from the HR system, transformed and feature selection was applied, to select features to be used as input.

In the following, we present the results from the process, as applied to one of the roles advertised to internal candidates, in the Investment and Financial function of the company (referred as Job A).

Data Collection. According to the job posting the requirements of this position were the following

- University degree in Finance/Economics
- Relevant work experience in finance
- Excellent knowledge of MS Office
- Very good knowledge of English, both oral and written.

There were 100 applicants, 57% females, and 43% males. Of the 100 applicants, 20 (20%) were qualified for further evaluation, by HR experts, and 80 (80%) were not. The following features were selected from the dataset and used as input from the algorithm: “Years in the Organization”, “Previous work experience”, “English knowledge”, “Education level”, “Education field”, “Professional Certification”. The output was the attribute “Short List”, which is a binary variable indicating whether an application will proceed to the next step. Taking into account the requirements of the job, the attributes “English knowledge level” and “Education level” were transformed into “English Knowledge Required” and “Education Level required”, respectively. The attributes used in the Naïve Bayes model construction, are depicted in Table 1.

Table 1: Attributes for model creation for Job A

Name	Description	Type	Values
Years in the Organization	Years in the Organization	Ordinal	NA, <1, 1-5, 5-10, 10-20, >20
Previous work experience	Previous work experience	Ordinal	NA, <1, 1-5, 5-10, 10-20, >20
Level	Seniority Level	Ordinal	1,2,3,4,5
English knowledge Required	Whether a candidate’s English knowledge satisfies the job requirement	Categorical	No Yes
Education level Required	Whether a candidate’s level of education satisfies the job requirement	Categorical	No Yes
Education field Required	Whether a candidate’s education field is among that is required in the job position	Categorical	No Yes
Profesional Certification	Whether a candidate has had Professional Certification	Categorical	No Yes
Short List	Selected for further Evaluation (output variable)	Categorical	No Yes

Next the training and test steps were carried out for the creation of the model. The available data were randomly divided into a set of training and test data into a ratio of 70:30. The method used for the data splitting was stratified random sampling. From the available data, 70 instances were randomly assigned to the training set (14 of the selected applicants / 56 not selected) and 30 to the testing set (6 of the selected applicants / 24 not selected).

The probabilities of every attribute-value pair with each value (Yes / No) of the target variable “Short List”, were computed from the training dataset. The likelihood values of Yes and No were calculated for each applicant, by employing the probabilities computed previously. For each event, the probability of “Yes” and “No” for the target variable “Short List” were calculated. The decision rule for classification on test data, was:

If Probability of Yes > Probability of No, then the prediction is Yes, and

If Probability of No > Probability of Yes, the prediction is No

Results and Discussion

The confusion matrix was generated for both the training and testing data (Table 2). In the matrix, we can see that the overall success rate of the model for both training and testing data is 80%, however, the model in testing data cannot predict the class “Yes”, which refers to applicants who are recommended for further evaluation by HR experts.

Table 2: Confusion matrix for Job A

Sample	Observed	Predicted		
		Short list -Yes	Short list -No	Percent Correct
Training	Short list -Yes	7	7	50.00%
	Short list -No	7	49	87.50%
	Overall Percent	20.00%	80.00%	80.00%
Test	Short list -Yes	0	6	0.00%
	Short list -No	0	24	100.00%
	Overall Percent	0.00%	100.00%	80.00%

The performance metrics of the algorithm for a series of job positions are also depicted in Table 3. Job A was the first one, and we can see the model's overall performance, measured by the metric accuracy, is acceptable, as it is over 60% for the majority of job positions. Its accuracy in predicting the negative class is also acceptable, however, its accuracy in predicting the positive class "Yes" is average to limited (metrics Precision, Recall and F1). This indicates that applicants who actually have been qualified for further evaluation have been excluded by the model.

Table 3: Model performance per Job position

	Job						
	A	**B**	**C**	**D**	**E**	**F**	**G**
Accuracy	80.00%	71.88%	75.00%	78.57%	92.31%	56.10%	68.42%
Precision	0.00%	44.44%	66.67%	66.67%	0.00%	56.25%	62.50%
Recall	0.00%	50.00%	25.00%	80.00%	0.00%	45.00%	62.50%
F1	-	47.06%	36.36%	72.73%	-	50.00%	62.50%

Low precision indicates that the model recommends candidates for further evaluation that actually are not qualified by company HR professionals (high False Positive recommendations). These wrong recommendations can impose costs for the company, such as extra time that HR experts have to spend on the wrong applicants. Low recall indicates that the model does exclude applicants who fit the job requirements and would have been considered qualified by HR professionals for further screening (high False Negative). This type of error can raise ethical issues to the Company and complaints by employees.

This study is limited by the following facts:

1. The scope of this decision support system is focused on the first screening phase of job applicants. As a result, the data used were limited mostly on personal information such as job experience, education, English language and professional certifications.
2. The available amount of data for training and testing the algorithm, for each job position, is limited.
3. The imbalanced classes, which reduce the accuracy of the algorithm, were present in almost all job positions except jobs F&G.

Although model's relatively low accuracy degree, it is extremely simple, explainable to non- experts and easy to implement. In future, data can be enhanced with additional features and include interview scores, personality test results, and cognitive ability test results, so it is expected that accuracy can be improved significantly. Further research would also explore other classifiers, such as decision trees, in order to evaluate and compare.

Conclusions

This paper focuses on automated decision-making in Human Resources and integration of automated decision-making in various functions of Human Resources Management Some key challenges include ethical and legal considerations, data protection, explain ability, and accountability issues. The question of fully automated or algorithmic augmented decision-making is addressed in the literature survey and it seems that a balanced collaboration of humans and AI is needed, for improved decision-making. In this context, we proposed a decision model based on Naïve Bayes classifier, and suggest that it can be utilized as a decision support system in the first phase of the employee selection process, one of the key functions of Human Resources Management.

We presented the key steps of the model, along with an application in a real job selection process, using real data originated from a large company in the financial sector. After appropriate transformation and coding, the model was trained and tested. Following that, the evaluation metrics were calculated and results indicate a relative decent level of accuracy, that although is not very high, it can be improved in future research, by enhancing feature selection and training process.

References

1. Garg S, Sinha S, Kar AK, Mani M. A review of machine learning applications in human resource management. International Journal of productivity and Performance Management, January; c2021. Retrieved from: https://doi.org/10.1108/IJPPM-08-2020-0427
2. Vassilopoulou J, Kyriakidou O, Özbilgin MF, Groutsis D. Scientism as illusion in HR algorithms: Towards a framework for algorithmic hygiene for bias proofing. Human Resource Management Journal, 2022 Jan 4, 1-15. DOI: 10.1111/1748-8583.12430
3. Meijerink J, Boons M, Keegan A, Marler J. Algorithmic human resource management: Synthesizing developments and cross-disciplinary insights on digital HRM. The International Journal of Human Resource Management. 2021 Jul 1;32(12):2545-2562. DOI: https://doi.org/10.1080/09585192.2021.1925326
4. Cheng MM, Hacket RD. A critical review of algorithms in HRM: Definition, theory, and practice, 2019, June 21. Retrieved from https://doi.org/10.1016/j.hrmr.2019.100698
5. Leicht Deobald U, Busch T, Schank C, Weibel A, Schafheitle S, Wildhaber I, Kasper G. The Challenges of Algorithm Based HR Decision Making for Personal Integrity. Journal of Business Ethics. 7 June 2019 Dec;160(2):377-392. Retrieved from https://doi.org/10.1007/s10551-019-04204-w
6. Noe RA, Hollenbeck JR, Gerhart B, Wright PM. Human Resource Management: Gaining a Competitive Advantage (10e Ed.). New York: McGraw-Hill Education; c2015.
7. Noe RA, Hollenbeck JR, Gerhart B, Wright PM. Fundamentals of Human Resource Management (6e Ed.). New York: McGraw-Hill Education, 2016.
8. Castellanos S. HR Departments Turn to AI-Enabled Recruiting in Race for Talent. The Wall Street Journal, March 14; c2019. Retrieved from https://www.wsj.com/articles/hr-departments-turn-to-ai-enabled-recruiting-in-race-for-talent-11552600459
9. Cappeli P, Tamble P, Yakubovich V. Artificial intelligence in human resources management: Challenges and a path forward, April 15 2019 Aug;61(4):15-42. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=32

63878
10. Pampouktsi P, Avdimiotis S, Maragoudakis M, Avlonitis M. Applied Machine Learning Techniques on Selection and Positioning of Human Resources in the Public Sector. Open Journal of Business and Management, 2021;9(2):536-556.
DOI: https://doi.org/10.4236/ojbm.2021.92030
11. Witten IH, Frank E, Hall MA. Data Mining Practical Machine Learning Tools and Techniques (3rd edition Ed.). Morgan Kaufmann; c2011.
12. Khairina D, Maharani S, Ramadiani, Hatta H. Decision Support System for Admission Selection and Positioning Human Resources by Using Naive Bayes Method. Advanced Science Letters. March 2017;23(3):2495-2497.
DOI: https://doi.org/10.1166/asl.2017.8653
13. Al-A'araji N, Al-Mamory S, Al-Shakarchi A. Constructing decision rules from naive bayes model for robust and low complexity classification. International Journal of Advances in Intelligent Informatics. 2021 Mar 1;7(1):76-88.
DOI: https://doi.org/10.26555/ijain.v7i1.578
14. Olson DL, Delen D. Advanced Data Mining Techniques. Berlin Heidelberg: Springer; c2008.