sustainability

MDPI

Article

An AI-Based Shortlisting Model for Sustainability of Human Resource Management

Erdinç Aydın * and Metin Turan

Faculty of Engineering, İstanbul Ticaret University, Istanbul 34840, Turkey
* Correspondence: erdinc.aydin@genom.com.tr

Abstract: The adoption of artificial intelligence in human resource management may help businesses and create a keen advantage in the market. With the help of artificial intelligence, most human resource duties can be completed efficiently and in a much shorter timeframe. For the sustainability of companies, it is essential to shorten the processes that are time-consuming and possible to automate. Especially in the recruitment process, artificial intelligence can ease short listings and much more. This study focuses on the adoption of artificial intelligence for recruitment and shortlisting as a human resource management operation. It is intended to remove noisy data from the resumes of applicants by using a minimum description length algorithm and to create a learning algorithm based on the support vector machine to choose the better candidates according to company culture and preferences. By creating shortlists for open positions, it is possible to improve the hiring process and cut the cost of the process. To the best of our knowledge, no studies in the research literature that focused on resume shares learning algorithms and performance evaluation results. This paper presents how the feature extraction algorithm fails while feature selection reduces successfully, and how the learning algorithm can create shortlisting.

Keywords: artificial intelligence; feature selection; hiring; human resource management; MDL; NLP; recruitment; support vector machine

check for updates

Citation: Aydın, E.; Turan, M. An AI-Based Shortlisting Model for Sustainability of Human Resource Management. *Sustainability* **2023**, *15*, 2737. https://doi.org/10.3390/su15032737

Academic Editor: Jun (Justin) Li

Received: 12 December 2022
Revised: 19 January 2023
Accepted: 31 January 2023
Published: 2 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Human resource management (HRM) has a long history of use. According to Jianhao, HRM was first invented in the 1960s and can be applied to anyone with a role in the company. Human resources (HR) also manages departments within the organization if they are directly or indirectly involved with employees [1]. HR has two primary purposes in an organization. The first is administrative jobs, and the other is employee management. The former is relatively easy because most of the rules are technically known. Still, even rule-based functions of HR should be reviewed and corrected if they create dissatisfaction among employees. This puts pressure on HRM to constantly monitor the employee. HR should focus on functions that are as important as possible to create a sustainable business rather than dealing with time-consuming processes.

People's lives have always been affected by new technological developments, and businesses have to keep up with developments. As a result, HRM has been reshaped by technological development. According to Stone, the principal goals of HRM are to gain and retain new talent, motivate employees, and select the best candidates. Technological developments have changed the functioning of HR processes, employee management, and job descriptions [2]. Of course, most HR process improvements have become possible due to the increase in data and knowledge. As García-Arroyo and Osca Segovia point out, information technologies enable more decentralized decision making so that more decision makers can contribute. This results in more advanced capabilities and better conversion of inputs (data) into outputs [3]. García-Arroyo and Osca also noted that data usage and big data have been transforming HR operations. They believe that the main challenges are

the acquisition, processing, and analysis of data in the interests of companies. However, there are still questions such as what size, how to organize, how to access, and how to distribute [4]. The complexity, size, and speed of the data flow allow machine learning algorithms to analyze and predict results based on data. According to Hong, current HRM systems can collect, query, and stats data, but it is required to use machine learning (ML) algorithms to be able to analyze the relationship between the data [5].

Keeping up with the latest developments increases their competitive advantage in the market and helps cut costs. Technological advancements not only automate and save time but also present a process that is very effective and fair. Any technology can only grow when it is at the right time. According to Haenlein and Kaplan, the rise of "Big Data" and advances in computing power have opened up new possibilities for AI [6]. There are still some obstacles and challenges regarding AI developments, such as training data shortage [7], regulations, and lack of standards [8]. However, it should also be suitable for the day and should be adoptable by employees, managers, and stakeholders. Nowadays, there are opportunities to adapt artificial intelligence (AI) and AI-based technologies, such as machine learning and natural language processing, to the different HRM phases. At the same time, ethical issues and transparency concerns have been started to discuss [9]. When automation is mentioned in every sector, there is a fear of unemployment. As Kuzior stated, every technological development shifts human labor and needed skills that are needed because of redesigned job descriptions. She stated that new technologies may create technological unemployment, which is considered short-term unemployment and can be overcome by proper education to gain the required specialized competencies [10]. In addition, Kuzior and Kwilinski say, in their article, that the social aspect of AI subcategories is not fully comprehended by every individual. Especially depending on age categories, the awareness of AI interaction changes. However, they noted that since the industry needs specialists in new technologies, it is crucial to get an education [11]. Moreover, there will be more challenges and more gray areas as AI takes shape in human life and HR in the future.

Berhil and others claim that machine-learning-based AI modernization can revolutionize HR departments at different levels. Recruitment, management, training, and benefit departments are examples that can use AI as a revolution in HR. They reported that AI can provide a huge benefit for candidate attraction and predict the candidate's added value for the company [12]. Bhardwaj and others wrote that the strength and growth power of organizations come from how they intelligently merge manpower and machinery at a minimal cost. They wrote that to remain competitive in the current market; companies should understand the capabilities of AI and redefine HR operations to benefit from artificial intelligence [13]. This project aims to realize the target by creating an algorithm that constantly learns to evaluate job seekers and create a shortlist of candidates to save time and cost. According to Ore and Sposato, candidate screening can take more than several weeks, but AI is not just solving that but also can eliminate unconscious human bias and enables HR to focus on attracting the best candidates to keep the competitive advantage for the company [14]. Yet, as Fan and others stated, since AI and ML algorithms are black-box solutions, it is possible that AI also may make an unethical or discriminatory decision as well [15]. In addition, since artificial intelligence can only be successful to the extent that it is fed from the data given to train, if there is a shortage of training data, it is possible that AI can make a biased selection which HR managers have done before. AI is only as strong as the given data so it can create the same result as the HR manager [7]. This project aims to solve both the problem of spending time to create a shortlist and to remove human biases by eliminating them before data selection by operator guidance (HR managers). Therefore, first of all, a data set is created for learning algorithms. Later, elements of the resumes, which are the features of learning algorithms, will be ready for the algorithms. Feature selection and feature extraction will illuminate which elements have been used so far and should be used for learning algorithms. As a final step, the performance evaluation of the learning algorithm will be examined. This process will prove that AI algorithms can mimic HR and have the ability to do more by constantly learning. There are few academic

studies that focus on personality tests; however, in this project, resume features will be the focus, and eliminating noisy features of resumes will be the subject. This study will include literature on the recruitment process and AI in the recruitment process, followed by data selection as machine learning algorithms and support vector machine (SVM) learning algorithms with the results on data.

Denise and Suzanne said that HRM is a major contributor to firms' competitive advantage. They also stated that the sustainability of HR can be achieved through the realization of a long-term people management approach, resulting in sustainable business performance and employee benefits [16]. The first step to employing a long-term committed employee is recruiting the right one for the position, which keeps them motivated. So, employee retention starts by picking the one who is right for the position [17]. Humans are the only part of a company that is not possible to imitate easily. Therefore, it is important to recruit the right employee as a resource for a company. It is possible to lose a candidate by putting them in a long and ambiguous selection process. In addition, if the hiring process is taking too long and the department is in urgent need of the workforce, they may need to hire an incompetent employee and then repeat the whole process. That is why the hiring stage needs to be completed as flawlessly as possible and as fastest as it can be. Since shortlisting is the most time-consuming part of that process and the most suitable process to automate, this project focuses on that subject.

There are few studies on the same topic, but they generally focus on personality tests and how it is possible to use these tests for candidate selection. Based on the information obtained from the interviews with HR managers responsible for recruitment, it was concluded that the personality test is not an important feature of the selection process. A recruiter stated that while personality tests are required prior to recruitment, they are not used to screen or prioritize. However, to our best knowledge, there is no AI project paper that is based on the resume features of applicants and shares the performance results of applied algorithms and models. This project was conducted recently with a 2022 dataset of applicants. The use of data and modeling were carried out, taking into account the latest recruitment standards of HR managers.

2. Background

2.1. Recruitment Process

Recruitment is one of the most important tasks of HR, because only by recruiting the most suitable employees for the right job can the company continue its business activities with high efficiency. According to Saad et al., in this digital age, the recruitment process has changed dramatically in innovative ways since sourcing the most suitable candidate for positions that are far more important to the company's competitive advantages. COVID-19 also pushed HRM to further digitize from conventional hiring [18]. According to Furtmuellera et al., curriculum vitae (CV) (also known by the name resume) includes the candidate's education, skills, strengths, achievements, and other job-related attributes. The pre-screening of a CV is a crucial step in recruitment selection since it has the ability to demonstrate the skills, motivation, capabilities, and potential value of applicants. it can be used to draw conclusions about the candidate [19]. Wilfred said that screening resumes is the hardest and most time-consuming part of the recruitment process. The larger the resume pool gets, the more difficult and time-consuming to screen it. In addition, it is always possible to miss the best fit for the openings when there are too many resumes to evaluate [20]. Furthermore, when there are so many resumes to consider, it is always possible to miss out on the one that best suits job postings. After reviewing applicants' CVs, if HR approves, they are added to the shortlist, and usually, the next phase begins once the shortlist is complete. As such, applicants generally have to wait for the shortlist to be complete, or is at least sufficient.

As Arbab and Mahdi mentioned in their article, the main goal of HRM is to gather competent people in the company, train personnel, make them perform at their best effort, and ensure that employees maintain their productive affiliation to benefit the organization.

However, the main differentiated value of HRM comes from the ability to acknowledge the personal skills of individuals and transfer the knowledge within the organization [21]. To realize that task, HR has imported many strategies and technologies from different domains. Psychometric tests, online analytical processing, and AI are examples of disciplines applied to HRM [22]. According to da Silva et al., HRM moves towards a digitalized age and digital technologies. This requires a staff member to understand and work with machines to complete their tasks. Training employees and requiring qualified personnel to work in a digitalized environment are also important. Therefore, the transformation of HRM to work with machines requires workforce changes, as well as in-house training and acquiring new digital-age skills for employees. They need to adequately analyze or understand AI sufficiently to use the collected data for the benefit of the organization [23]. According to Stone et al., thanks to technological developments, HR is now not just responsible for mediating between the organization and individuals, but also for removing the distance barrier and enabling a remote working environment. This requires a global recruitment process to obtain global talent [2]. The increase in the number and diversity of applicants requires both online and automated HR processes. Therefore, the need for data management is a critical operation for an organization. This is one reason to adopt AI systems for HR management. AI can replace most routine HR tasks with minimal human intervention. There are also many benefits to adopting AI, which enable the process to be completed in less time with increased accuracy [13].

2.2. Recruitment Process with AI

Recruitment is one of the most commonly used AI-embedded processes. AI assists in various tasks of the process and improves experience while saving time. It scans high-volume resumes and evaluates them automatically to create a shortlist at no time [24]. This shortened time is not just for cost-cutting, but also improves companies' image and helps to protect competitive advantage over the market. According to Alavuo, a candidate-centric recruitment process is the key to becoming appreciated by the candidates and gaining a competitive advantage over competitors. To achieve this goal, the recruitment process should be easier, and more importantly, clear communication should be maintained between applicants and the recruitment process. If a rejected applicant is still interested in the company for the future and recommends the company to its network, the recruitment process has been run successfully [25]. Hoang argued that traditional hiring management lags behind new areas' talent wars. Winning this war is especially important in competitive industries. That is why Hoang suggests strategic long-term talent solutions for recruitment departments [26]. AI solutions can solve many problems using techniques such as chatbots, natural language processing (NLP), and deep learning algorithms.

2.3. Some AI-Related Research in HRM

There have been some attempts at AI adaptation for HR. For example, Ismail et al. concluded that only large companies are adopting AI in recruitment, and to what extent they are using it is still unclear. So, it is not clear which techniques or methods are being used in AI for recruitment currently [27]. Mihuandayani et al. created a machine learning algorithm using SVM for personality profiling of candidates on social media platforms. They obtained 64.5% classifying accuracy for five classes (personality types) [28]. As the ROC, AUC, and confusion matrix were not given, 64.5% accuracy is not sufficient to conclude whether the model was successful or not. More importantly, using personality types is not a sufficient selection criterion for recruitment. Another research was made by Qiangwei and colleagues. Data (about HR features) were collected by questionnaire survey. Affinity propagation (AP) and SVM were used for creating a new feature selection algorithm. They reduced 24 features with the new model to 12 features and increased accuracy from 84.98% to 85.84% [29]. On the other hand, Yung-Ming executed research for recruitment using the SVM learning algorithm. Data were collected through personality surveys of current employees and candidates. Even though conventional thinking suggests

that skill score is of greater importance than personality, their research finds that personality is a key factor in evaluating an employee for a position [30].

2.4. Opportunities in AI Applications for HRM

2.4.1. Chatbots and NLP in Recruitment

Chatbots are AI applications widely used in HR management projects. For this paper, NLP and chatbot applications will not be the subject, but they might be considered in future studies since they help the recruiter and create a candidate-centric hiring process. As Soutar's study shows, e-commerce, healthcare, and insurance sectors have been reshaped by chatbot adoption, but the recruitment sector has less than 30 percent of usage. As Mnasri explained, rule-based chatbots do not require AI or NLP but follow a pre-programmed pattern. Data-driven chatbots can create long and complex conversations. Therefore, it is possible to collect much more output from data-driven chatbots owing to the NLP [31]. Ambiguous parses of words and ungrammatical spoken prose problems which require deeper analysis and machine learning techniques to solve the meaning of the texture. The algorithms try to solve the most probable outcome (meaning) of the given inputs because it is not possible to detect the meaning with preprogrammed codes [32,33]. Companies believe that chatbots can have direct advantages over email commutation, which is the preferred communication with applicants now [34]. Chatbots are an advanced field of NLP applications that facilitate communication-based automation without the need for an operator. Chatbots create prompt communication between candidates and the HR department [35]. According to Nawaz and Gomes, chatbots can answer questions about candidates up front, such as salary range, leave facilities, FAQs, and incentives about the workspace [36]. This can automate some of the HR workloads and answer candidates' questions 24/7 and instantly. Anitha concluded that thanks to chatbots, employers are able to save 74% of recruitment marketing efforts, which include time and budget expenses. In addition, chatbots have a positive effect on candidate engagement and enable easier handling of onboard activities [37]. According to Koivunen et al., a chatbot can easily obtain inputs such as address, education level, and experience through low-level chatbot applications. However, a high-end chatbot (which requires AI) can handle almost any candidate interview. In this study, it was stated that an interview chatbot, which requires AI to understand sentences, could shortlist 316 candidates to 12 candidates [38].

2.4.2. Data Mining for AI

According to Shehu and Saeed, data mining is an information technology domain that combines statistics, machine learning, and artificial intelligence [39]. Data mining is the process of discovering meaningful and useful information from previously unknown or hidden patterns in a dataset [40]. In AI, data mining techniques can be used to obtain the results of four main functions: association, clustering, classification, and prediction. They argued that, by using data mining, it is possible to extract value from the curriculum vitae of applicants, which can be used in AI systems [41]. According to Singh and Kumar, Data mining is possible using various techniques on a given dataset, such as decision tree, k-nearest neighbor, naïve Bayes, support vector machines (SVMs), neural networks, and decision support models. Given that classification problems pre-define the labels (classes) for a given dataset, according to Pah and Utama, the flow of the mining algorithm consists of five main functions, including data collection, data preparation, data mining, and data mining evaluation (determining the best model that results in the highest accuracy). Then, it is possible to obtain the classification rules based on the created model if it was sufficiently successful and passed the performance evaluation [42]. Using data mining anomaly detection, association rule learning, clustering, classification, regression, and summarization can be achieved [43]. However, for the recruitment process, classification is a problem that must be solved by AI. According to Pah and Utama, as organizations grow, an increasing number of criteria (features of machine learning) should be considered

when choosing the desired staff as employees. Therefore, it requires a longer time, and data mining algorithms to save time and cut expenses [42].

3. Methodology

The aim of this study was to create a shortlist from the CVs of the candidates. Since NLP is not the subject of this study, CV data were recorded in the relational database system (RDBMS) as in medium- and large-sized companies. In other words, CVs are not read from papers but stored digitally in database tables, which makes data preparation easier. As mentioned previously, there are at least four stages of machine learning and AI. The first stage involves preparing the data, creating the model, and analyzing the performance of the model [42]. According to Willemink et al., the data must first be acquired. Thus, accessing and querying the data is the first stage of data preparation for all machine learning or artificial intelligence projects. Then, de-identification can be applied to the dataset owing to privacy concerns. Labeling the data, formatting the data, and storing the new data must also be realized before using machine learning methods [44]. After all of the preparations depending on the model data can be cleaned, missing data can be filled in. Feature selection or extraction is one of the most important stages in data preparation. According to Soibelman and Kim, the feature selection of a dataset directly affects the performance of the model because some of the features (inputs) do not have a relationship with the desired output [45]. As Kumar and Minz mentioned, feature selection may be a difficult task because of the high number of irrelevant or redundant features or noisy data, if it is not possible to know the features beforehand [46]. According to Chandrashekar and Sahin, creating a new subset with feature selection can be applied using a genetic algorithm, sequential floating forward selection, sequential feature selection, SVM classifier, or unsupervised learning algorithms [47].

3.1. Minimum Description Length

Grünwald described the minimum description length (MDL) principle, which is a relatively recent method, as a model selection principle for inductive inference, which provides a solution for selection problems while avoiding overfitting problems [48]. As Grünwald states, the purpose of the MDL is to compress the meaningful data, which is referred to as 'Regularity'. Then, it can compress the meaningful features as best as possible by combining the existing features. It can be used as a simple MDL or can be transformed to higher (more complex) polynomial degrees to describe information about the data [49]. As Grünwald and Roos state, most MDL algorithms are related to Bayesian model selection [50]. In addition, Li and [51] stated that MDL minimizes the sum of the following:

- The length, in bits, of the description of the theory; and
- The length, in bits, of data when encoded with the help of the theory.

According to Quinlan, if the classifier labels the target class by making the prediction (described as categorical), then the MDL increases the positive and negative errors. Therefore, the algorithm performed poorly and required a biased strategy [52].

3.2. Support Vector Machine

According to Vishwanathan and Murty, the SVM is one of the most significant learning algorithms for classification problems. It is possible to classify a given dataset using linear or non-linear separation surfaces [53]. As Mahesh stated, the SVM is a supervised learning algorithm, so it is necessary to label the target class before handling it with SVM. If linear classification or regression analysis is not successful, it is possible to move inputs to higher dimensions using kernel tricks [54]. Orrù and Francesco et al. stated that the SVM tries to create a maximum margin, which is the maximum distance between the plane and the nearest data [55]. According to Byvatov et al., an SVM works by transforming sample data to a higher dimension than the original data. Subsequently, it attempts to find the maximal margin separating the classes of data. If it is not possible to separate classes, even

with much higher dimensions, slack variables should be used to formalize the SVM [56]. According to Evgeniou and Pontil, the slack variable represents the error at the point [57]. A graphical representation of the SVM is shown in Figure 1. Since it is not possible to differentiate two different classes (colors of circles) perfectly just with a linear hyperplane, there is a slack error value.

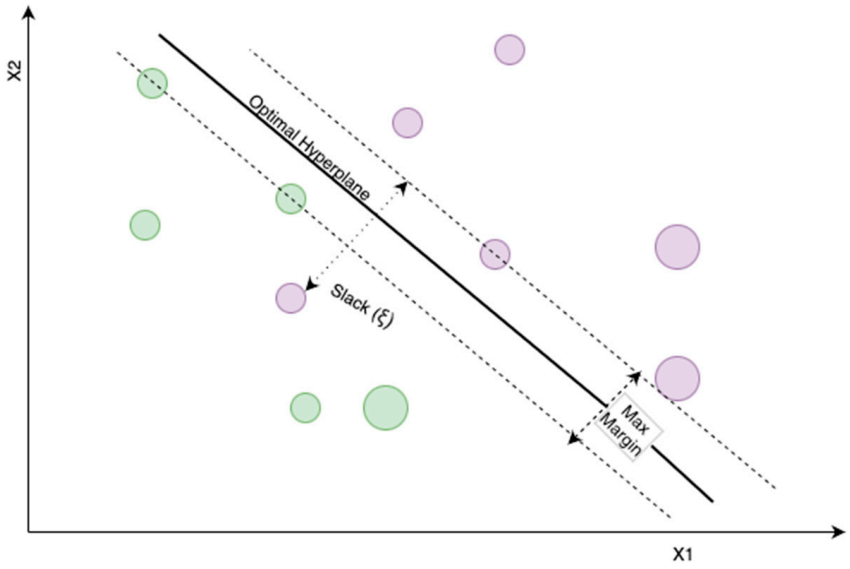


Figure 1. SVM algorithm.

3.3. Data

The first step of AI is data acquisition and data access. Because for a learning algorithm, data must be real or realistic, so AI can learn as it should. As Prasanna and others stated, in HRM, it is not possible to collect every aspect to find a good employee, and not all data that are collected are relevant to employee acquisition. Moreover, there is no standardization for hiring, and every company has its own organizational culture [7]. Therefore, it is possible to obtain different features for every dataset that is important or relevant. In addition, some features can be regarded as company culture. In this project, the age and sex of employees were rejected before feature selection because of HR feedback for the hiring process. Even if human bias is affected by past hirings, these features are no longer welcomed for a hiring process. A real job description was posted on a job-seeking website in Turkey to obtain applicant resumes. An information technology (IT) position was created to employ data analysis with remote working opportunities. It is expressed that the English language is mandatory, and it is necessary to have at least two years of work experience. In addition, candidates are asked to have strong written and verbal communication skills and to be careful and disciplined. In total, 140 resumes were collected from the website. In this project, the resume information of real data was used, but for privacy and legal regulations, all information was de-identified (anonymization of data). In Turkey, the personal data protection law clearly states that it is necessary to apply for de-identification or that permission must be signed by the applicants. After obtaining and storing the data, conversions must be performed before creating a valid subset. The project steps are illustrated in Figure 2. The aim was to use the SVM before creating the model, which is necessary to convert categorical features into numerical representations. Thus, some features, such as graduation status, military status, city, district, blood type, and driver's license class, were converted into numerical representatives. It is possible to deal with missing and redundant values with feature selection algorithms; however, in this project, they were handled as a step of data preparation before running any algorithm.

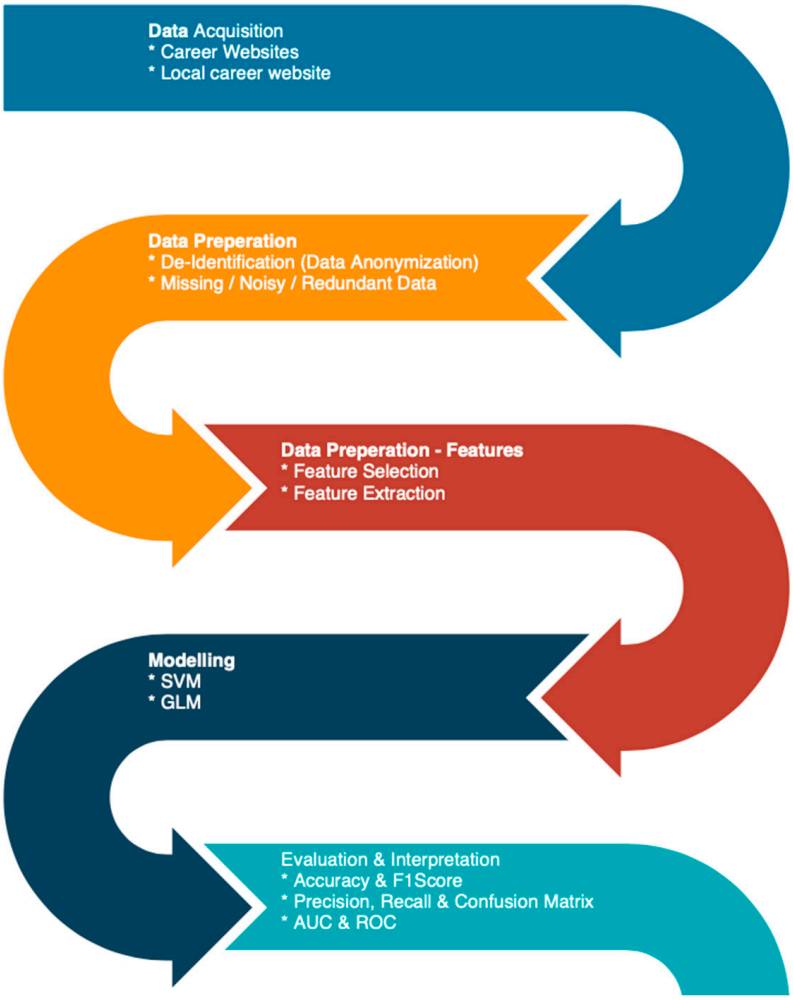
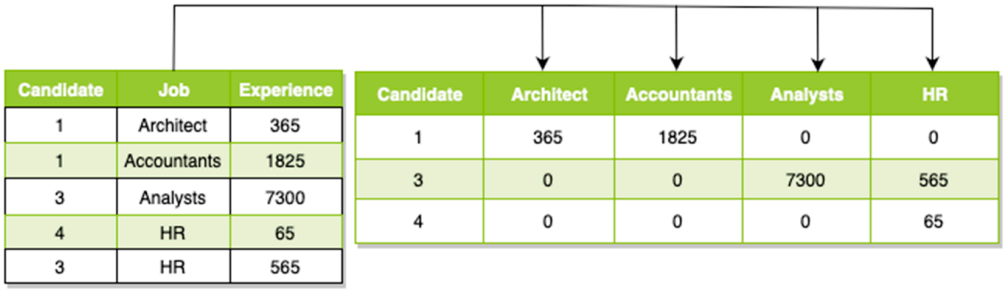


Figure 2. All the steps of AI learning.

Only 100 data points (applicants) were used in training and testing for both the data selection algorithms and training algorithms. All the data of resumes were kept in the database (features of learning algorithm), but phone numbers, names, surnames, exact addresses, and other information that can be used to identify the person were omitted or anonymized before saving in the local database. Because all data were stored in a relational database and all the models only work on a single table, de-normalization must take place before modeling. In the case of redundancy being created, it must be handled as a data preparation step.

Furthermore, additional data preparation is needed because some information is stored in rows, not in columns, since normalization in RDBMS requires it. For example, past job experiences of a candidate (might be multiple on a CV) can be joined to a candidate as an NxN relationship. However, every candidate must have only one row in the dataset for the machine learning algorithm. That is why every past job experience (Analysis, Software Developer, Manager, etc.) needs to be a separate feature rather than a value of features. For this example, it is necessary to multiply years by 12 and add months as the value of job positions. This multiplies the feature numbers and creates noisy data, as shown in Figure 3.



Candidate	Job	Experience
1	Architect	365
1	Accountants	1825
3	Analysts	7300
4	HR	65
3	HR	565

Candidate	Architect	Accountants	Analysts	HR
1	365	1825	0	0
3	0	0	7300	565
4	0	0	0	65

Figure 3. Converting database column values (rows) to columns (features of learning algorithm), which created 18 additional columns to use as features.

3.4. Feature Selection

For this project, features that can identify the candidate were removed or obscured before saving the data. In addition, some characteristics such as gender or age were eliminated by an operator since it was accepted as unethical to use in the hiring process in the new HR approach. This was not taken into account in previous studies, which followed the data as they were. However, HRM concepts are changing, and it is not always enough for AI to replicate recruitment specialists in the past of the company. The last step in data preparation is data selection. In this study, the MDL-based model was used. A total of 107 attributes were constructed, but 85 remained after the data were cleaned. Cleaning involves removing all NULL (empty) attributes, attributes that cannot be categorized, and non-numeric attributes. For example, since the academic title attribute is empty in all candidates, it was cleared as data preparation, so it will not be used in the learning algorithm. In addition, "explanations" was removed from the feature since it is not possible to categorize, and NLP is not introduced in the project. The "Attribute Importance" function from the Oracle database uses MDL as its default. The MDL algorithm successfully removed noisy features from 85 to 6. Some features, such as candidate addresses (city or even country), were removed by the algorithm because it is an online job opportunity. Surprisingly, however, educational status and total experience were not also selected. Auto data preparation features were disabled from the attribute importance function to observe the real results of the MDL. Ranks show the important attributes of the job (Figure 4).

variable	importance	rank
BD_10059	0.1855	1
BYD_1	0.1563	2
BD_10021	0.0223	3
BSS_26	0.0162	4
B_CALISMA_DURUMU	0.0112	5
BD_GECICI_TABLO_EVT__	0.0109	6

Figure 4. Results of 7 highest rank of feature selection with MDL algorithm.

- BD_10059: Indicates that they have experience with the "Analysis" position.
- BY_1: Language knowledge "English" is important.
- BD_10021: They have past experience as "Software Developer".
- BSS_26: Has the knowledge for "Analysis".
- B_CALISMA_DURUMU: If they are working currently or not.
- BD_GECICI_TABLO_EVT__: Indicates that a candidate has an experience in any job.

3.5. Feature Extraction

Before executing the learning algorithm, it is often necessary to distinguish between relevant and irrelevant data. Consequently, getting rid of irrelevant data improves the performance of the learning algorithm. Feature extraction is to find a highly representative set of features from originals. Unlike feature selection, feature extraction attempts to find a significant meaning with a less dimensional representation of the original [58]. Removing irrelevant features increases the performance of the learning model. However, a drawback of feature extraction is the linear combination of features, which is generally not interpretable, so it is not always possible to calculate how much original data has been lost [59]. According to Zena and Duncan, principal component analysis (PCA) is one of the widely used techniques, which produces the covariance matrix and its eigenvectors (each representing some proportion of variance in the data) [60]. On the other hand, Abdi and Lynne expressed that PCA is actually the oldest multivariate technique, dating back to 1901 or even 1829. In addition, they stated that the goals of PCA are [61]:

- Extract the most important information from the original dataset.
- Compress the dataset keeping only this important information.
- Simplify dataset annotation.

To test the results of the PCA algorithm, the generalized linear model (GLM) was used as the learning algorithm instead of SVM. Of the data, 60% was used to train the GLM algorithm, and 40% of the dataset was used to test performance. After data preparation, 98 features of the resume remained as original dimensions of the learning algorithm. The PCA algorithm transformed the original 98 features into 25 new features. In addition, the percentage of the first three new features in recognizing the entire dataset, respectively, is as follows: 60.19%, 22.15%, and 8.68%. When all 98 features were tested (with GLM) without applying the suggested features of feature extraction, the accuracy was 65.0%. However, if feature extraction was applied using the PCA algorithm, the accuracy decreased to 37.5%. That is why feature extraction was unsuccessful for the project, and it will be ignored for the rest of the paper.

3.6. Machine Learning Model and Experiments

SVM has been used as the learning algorithm. By using the SVM, it was successful in classifying whether the candidates' resumes are suitable for shortlisting. For learning algorithm training, different subsets of features were tested to achieve the best result for the job position. Of the total samples, 80% were used as training data, and the remaining 20% were used as test data. If 70–30% was used, then the algorithm reacted the worst and only found five attributes as important. The SVM algorithm was used with all of the features and MDL-recommended features with linear kernel first.

The data were separated by 60% to 40% to fill the SVM as the dataset. If no feature selection was applied, the accuracy of the model was found 62.5%. The initial parameter settings are listed in Table 1.

Table 1. No feature selection—linear SVM.

Model	Setting	Value
No Feature Selection	Value of complexity factor	10
No Feature Selection	Number of Iteration	30
No Feature Selection	SVM Solver	SVMS_SOLVER_IPM
No Feature Selection	Convergence Tolerance	0.0001
No Feature Selection	Kernel	SVMS_LINEAR
No Feature Selection	Number of Features	82

Because many noisy features were maintained, this result seems normal. If only the features that the MDL algorithm discovers as significant are used, then with linear, kernel accuracy is 0.78. The initial parameter settings are presented in Table 2.

Table 2. Linear SVM with MDL selection.

Model	Setting	Value
No Feature Selection	Value of complexity factor	10
No Feature Selection	Number of Iteration	30
No Feature Selection	SVM Solver	SVMS_SOLVER_IPM
No Feature Selection	Convergence Tolerance	0.0001
No Feature Selection	Kernel	SVMS_LINEAR
No Feature Selection	Number of Features	6

Finally, a Gaussian kernel was used. The tolerance was set to 0.001 (the initial parameter settings are presented in Table 3), and the accuracy was 0.923. Even though this result shows satisfying accuracy, it is not sufficient to evaluate and accept the model by just looking at accuracy.

Table 3. Gaussian SVM with MDL feature selection.

Model	Setting	Value
No Feature Selection	SVM Solver	SVMS_SOLVER_IPM
No Feature Selection	Number of Iteration	11
No Feature Selection	Value of Standard Deviation	1.7320
No Feature Selection	Convergence Tolerance	0.001
No Feature Selection	Kernel	Gaussian
No Feature Selection	Number of Features	6

After observing every measure of ML algorithms to evaluate the model, Gaussian kernel and a tolerance value of 0.001 was found to be the best one. Although some projects require high recall, such as medical problems, this problem does not force obtaining a high value. However, the recall of the project was 0.875, which is sufficient to accept the model. The distribution of the prediction graph can be found below (Figure 5).

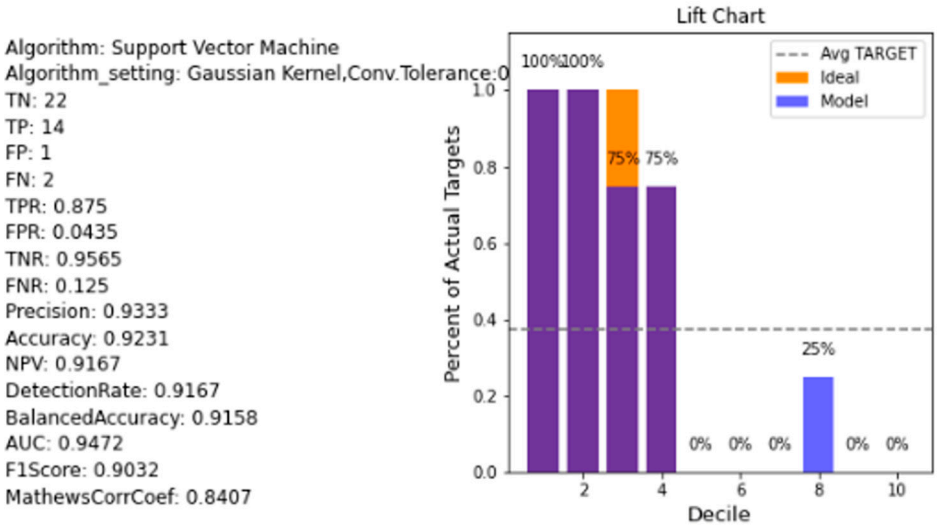


Figure 5. Results of SVM with Gaussian kernel and 0.001 convergence tolerance.

Since 0.5 probability separates two classes, true negative (TN) predictions are higher than others, and true-positive (TP) predictions are also high. According to Park et al., a receiver operating characteristic (ROC) curve is the sensitivity and specificity of TP/positive values over TN/negative values. This can also be defined as the TP and false-positive (FP) rates. More importantly, the area under the roc curve (AUC) was considered a measure of the overall performance of the training model [62]. Hoo et al. stated that the linear cutoff

line/reference line (X = Y) represents the rate of false-positive values, which is the same as the rate of true-positive values. Thus, it is necessary to create a model with the ROC graph above the reference line to be considered successful. More importantly, AUC is a global measure of the ability to evaluate a model's ability to distinguish binary classes [63]. The SVM model in the project resulted in 0.9472. AUC value shows that the model can separate applicants as important and create a shortlist for hiring. The AUC and ROC curves are shown in Figure 6.

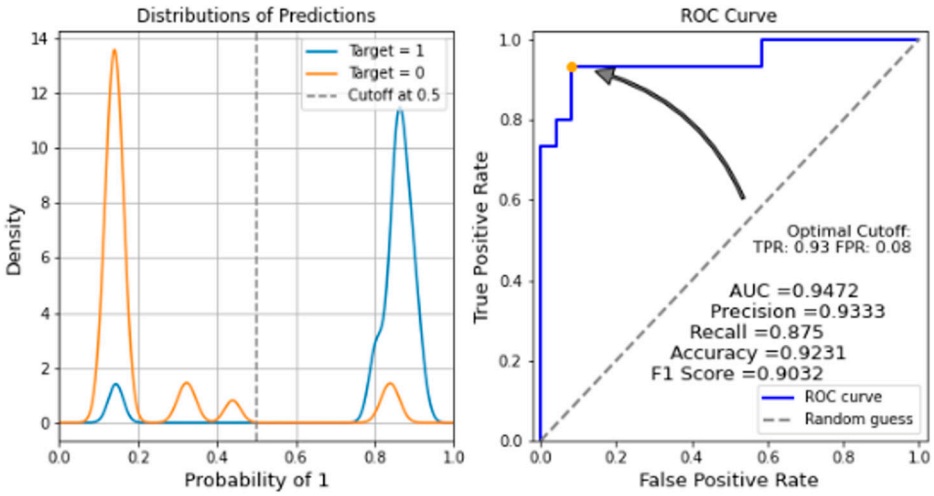


Figure 6. ROC and AUC results of SVM with Gaussian kernel and 0.001 convergence tolerance.

4. Conclusions and Discussion

Employees are inimitable assets that enable distinct competitive advantage. Therefore, recruitment is one of the most important tasks of HRM. However, as the recruitment process is a time-consuming process, it can exhaust applicants and create dissatisfaction. Companies try to ease the recruitment process as much as possible and simultaneously cut costs. If candidates are satisfied with the process whether they are hired or not, they will keep applying for future positions and recommend the company, which is important for future talent acquisition and sustainability of the workforce for a company. Artificial intelligence created a huge opportunity to create a shortlist from the resume information of the candidates. The features depend on both company culture and job position advertisements. Therefore, the feature selection must be applied before the learning model. Unfortunately, the feature extraction algorithm, which was an unsupervised method, was not successful to use in the study. In addition, there were no optimal settings for the learning model. Depending on the data, the SVM model can produce different results for each option such as kernel, tolerance, or iteration values. The SVM model successfully selected applicants for the shortlist, although the data were not sufficiently large. As the results are evaluated in the confusion matrix, the numbers of correctly labeled CVs are (TN + TP) 36 over 39 (all the test data). Including the training, both data selection and learning algorithm only took 3 s, which is an important advantage over humans since successfully labeled (shortlisted) data are 0.92%. Since in data selection, the features are clearly stated unlike black-box solutions, human biases also can be recognized. For this study, every feature was related to the position, but for a different dataset, it is possible to detect how HRM is creating a shortlist. Therefore, it will be possible to detect if race, age, and sex of applicants or graduated schools are the inputs of the shortlisting. This can be a game changer for companies to maintain their sustainability and competitive advantage. As mentioned earlier, hiring the most suitable candidate for the current position is the key to keeping the company operating smoothly. Every new hire is a long process and requires a new onboarding and learning phase for positions. This study proves that AI can be used to reduce workforce requirements and time costs while producing shortlisting

using candidates' resumes. It was also discovered that many of the resume features can be considered noisy values and need to be taken care of before using the learning algorithm. It is possible to detect which part of the resumes of candidates is considered while creating the shortlisting by HR previously. Moreover, the order of importance can be discovered between the characteristics of the resume with the data selection algorithm. The resume of the candidate can be considered much more effective and valid for a recruitment process instead of using personality tests. For future projects, the candidate database should be larger. Feature extraction algorithms should be examined for efficiency. Finally, NLP AI should be integrated into the model because, in real life, a resume is not sufficient enough. A letter of intent and a social network background check can be combined with the results of the resume inference.

Author Contributions: Writing—original draft, E.A.; Writing—review & editing, M.T. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by Genom Information Technologies.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jainh, M. A Comparison of Nationalized and Private Banks' Strategic Human Resource Management Practices. *J. Contemp. Issues Bus. Gov.* **2021**, *27*, 709–713.
2. Stone, D.L.; Deadrick, D.L.; Lukaszewski, K.M.; Johnson, R. The Influence of Technology on the Future of Human Resource Management. *Hum. Resour. Manag. Rev.* **2015**, *25*, 216–231. [CrossRef]
3. Marler, J.H.; Parry, E. Human resource management, strategic involvement and e-HRM technology. *Int. J. Hum. Resour. Manag.* **2016**, *27*, 2233–2253. [CrossRef]
4. García-Arroyo, J.; Osca Segovia, A. Big data contributions to human resource management: A systematic review. *Int. J. Hum. Resour. Manag.* **2021**, *32*, 4337–4362. [CrossRef]
5. Hong, Z. Research on human resource recommendation algorithm based on machine learning. *Sci. Program.* **2021**, *2021*, 8387277.
6. Michael, H.; Kaplan, A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *Calif. Manag. Rev.* **2019**, *61*, 5–14.
7. Tambe, P.; Cappelli, P.; Yakubovich, V. Artificial intelligence in human resources management: Challenges and a path forward. *Calif. Manag. Rev.* **2019**, *61*, 15–42. [CrossRef]
8. Tekkeşin, A.İ. Artificial Intelligence in Healthcare: Past, Present and Future. *Anatol. J. Cardiol.* **2019**, *22*, 8–9.
9. Buiten, M.C. Towards intelligent regulation of artificial intelligence. *Eur. J. Risk Regul.* **2019**, *10*, 41–59. [CrossRef]
10. Kuzior, A. Technological Unemployment in The Perspective of Industry 4.0 Development. *Virtual Econ.* **2022**, *5*, 7–23. [CrossRef]
11. Kuzior, A.; Kwilinski, A. Cognitive Technologies and Artificial Intelligence in Social Perception. *Manag. Syst. Prod. Eng.* **2022**, *30*, 109–115. [CrossRef]
12. Berhil, S.; Benlahmar, H.; Labani, N. A review paper on artificial intelligence at the service of human resources management. *Indones. J. Electr. Eng. Comput. Sci.* **2020**, *18*, 32–40. [CrossRef]
13. Bhardwaj, G.; Singh, S.V.; Kumar, V. An Empirical Study of Artificial Intelligence and its Impact on Human Resource Functions. In Proceedings of the 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 9–10 January 2020; pp. 47–51.
14. Ore, O.; Sposato, M. Opportunities and risks of artificial intelligence in recruitment and selection. *Int. J. Organ. Anal.* **2021**, *30*, 1771–1782. [CrossRef]
15. Li, F.; Ruijs, N.; Lu, Y. Ethics & AI: A Systematic Review on Ethical Concerns and Related Strategies for Designing with AI in Healthcare. *AI* **2023**, *4*, 28–53.
16. Jepsen, D.M.; Grob, S. Sustainability in recruitment and selection: Building a framework of practices. *J. Educ. Sustain. Dev.* **2015**, *9*, 160–178. [CrossRef]
17. Bauer, T.N.; Erdogan, B.; Taylor, S. *Creating and Maintaining Environmentally Sustainable Organizations: Recruitment and Onboarding*; Jossey-Bass/Wiley: New York, NY, USA, 2012.
18. Saad, M.F.M.; Nugro, A.W.L.; Thinakaran, R.; Baijed, M. A Review of Artificial Intelligence Based Platformin Human Resource Recruitment Process. In Proceedings of the 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Kedah, Malaysia, 1–3 December 2021.

19. Furtmuellera, E.; Wilderom, C.; Tate, M. Managing recruitment and selection in the digital age: E-HRM and resumes. *Hum. Syst. Manag.* **2021**, *30*, 243–259. [CrossRef]
20. Wilfred, D. AI in recruitment. *NHRD Netw. J.* **2018**, *11*, 15–18. [CrossRef]
21. Arbab, A.M.; Mahdi, M.O.S. Human resources management practices and organizational excellence in public orgnizations. *Pol. J. Manag. Stud.* **2018**, *18*, 9–21. [CrossRef]
22. Strohmeier, S.; Piazza, F. Artificial intelligence techniques in human resource management—A conceptual exploration. In *Intelligent Techniques in Engineering Management*; Springer: Cham, Switzerland, 2015; pp. 149–172.
23. Da Silva, L.B.P.; Soltovski, R.; Pontes, J.; Treinta, F.T.; Leitão, P.; Mosconi, E.; Yoshino, R.T. Human resources management 4.0: Literature review and trends. *Comput. Ind. Eng.* **2022**, *168*, 108111. [CrossRef]
24. Merlin, R.P.; Jayam, R. Artificial Intelligence in Human Resource Management. *Int. J. Pure Appl. Math.* **2018**, *119*, 1891–1896.
25. Alavuo, N.H. Modern Recruitment Process as a Competitive Advantage in Talent Acquisition: A Recruiter's Playbook. Master's Thesis, Haaga-Helia University of Applied Sciences, Helsinki, Finland, 2020.
26. Hoang, L. From Customer Journey to Candidate Journey: Applying Marketing Principles to Build a Winning Hiring Culture. Master's Thesis, Metropolia University of Applied Sciences, Helsinki, Finland, 2018.
27. Al-Alawi, A.I.; Naureen, M.; AlAlawi, E.I.; Al-Hadad, A.A.N. The Role of Artificial Intelligence in Recruitment Process Decision-Making. In Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 7–8 December 2021; pp. 197–203.
28. Utami, E.; Luthfi, E.T. Profiling Analysis Based on Social Media for Prospective Employees Recruitment Using SVM and Chi-Square. *J. Phys. Conf. Ser.* **2018**, *1140*, 012043.
29. Wang, Q.L.; Li, B.; Hu, J. Feature selection for human resource selection based on affinity propagation and SVM sensitivity analysis. In Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India, 9–11 December 2009.
30. Li, Y.-M.; Lai, C.-Y.; Kao, C.-P. Building a qualitative recruitment system via SVM with MCDM approach. *Appl. Intell.* **2011**, *35*, 75–88. [CrossRef]
31. Mnasri, M. Recent advances in conversational NLP: Towards the standardization of Chatbot building. *arXiv* **2019**, arXiv:1903.09025.
32. Chowdhury, G.G. Natural language processing. In *Fundamentals of Artificial Intelligence*; Springer: New Delhi, India, 2020; pp. 603–649.
33. Nadkarni, P.M.; Ohno-Machado, L.; Chapman, W.W. Natural language processing: An introduction. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 544–551. [CrossRef] [PubMed]
34. Soutar, K. How Chatbots Can Be Used to Re-Engage with Applicants during Recruitment. Master's Thesis, Aalto University, Espoo, Finland, 2019.
35. Majumder, S.; Mondal, A. Are chatbots really useful for human resource management? *Int. J. Speech Technol.* **2021**, *24*, 969–977. [CrossRef]
36. Nawaz, N.; Gomes, A.M. Artificial intelligence chatbots are new recruiters. *(IJACSA) Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*. [CrossRef]
37. Anitha, K.; Shanthi, V. A Study on Intervention of Chatbots in Recruitment. In *Innovations in Information and Communication Technologies (IICT-2020)*; Springer: Cham, Switzerland, 2021; pp. 67–74.
38. Koivunen, S.; Ala-Luopa, S.; Olsson, T.; Haapakorpi, A. The March of Chatbots into Recruitment: Recruiters' Experiences, Expectations, and Design Opportunities. *Comput. Support. Coop. Work. (CSCW)* **2022**, *31*, 1–30. [CrossRef]
39. Shehu, M.A.; Saeed, F. An adaptive personnel selection model for recruitment using domain-driven data mining. *J. Theor. Appl. Inf. Technol.* **2016**, *91*, 117–130.
40. Singh, S.; Kumar, V. Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques. *Int. J. Sci. Eng. Comput. Technol.* **2013**, *3*, 31–37.
41. Hmoud, B.; Laszlo, V. Will artificial intelligence take over human resources recruitment and selection. *Netw. Intell. Stud.* **2019**, *7*, 21–30.
42. Pah, C.E.A.; Utama, D.N. Decision Support Model for Employee Recruitment Using Data Mining Classification. *Int. J. Emerg. Trends Eng. Res.* **2020**, *8*, 1511–1516.
43. Joshi, A.P.; Panchal, R.K. Data mining for staff recruitment in education system using WEKA. *Int. J. Res. Comput. Sci. Manag.* **2014**, *2*, 1–4.
44. Willemink, M.J.; Koszek, W.A.; Hardell, C.; Wu, J.; Fleischmann, D.; Harvey, H.; Folio, L.R.; Summers, R.M.; Rubin, D.L.; Lungren, M.P. Preparing medical imaging data for machine learning. *Radiology* **2020**, *295*, 4–15. [CrossRef] [PubMed]
45. Soibelman, L.; Kim, H. Generating construction knowledge with knowledge discovery in databases. In Proceedings of the Eighth International Conference on Computing in Civil and Building Engineering (ICCCBE-VIII), Stanford, CA, USA, 14–16 August 2000; pp. 906–913.
46. Kumar, V.; Minz, S. Feature Selection: A literature Review. *SmartCR* **2014**, *4*, 211–229. [CrossRef]
47. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]
48. Grünwald, P.D. *The Minimum Description Length Principle*; MIT Press: Cambridge, MA, USA, 2017.

49. Grünwald, P. A Tutorial Introduction to the Minimum Description Length Principle. In *Advances in Minimum Description Length: Theory and Applications*; MIT Press: Cambridge, MA, USA; London, UK, 2005; Volume 5, pp. 1–80.
50. Grünwald, P.; Roos, T. Minimum description length revisited. *Int. J. Math. Ind.* **2019**, *11*, 1930001. [CrossRef]
51. Li, M.; Vitányi, P.M. Computational machine learning in theory and praxis. In *Computer Science Today*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 518–535.
52. Quinlan, J.R. MDL and categorical theories (continued). In *Machine Learning Proceedings 1995*; Elsevier: Amsterdam, The Netherlands, 1995; pp. 464–470.
53. Vishwanathan, S.V.M.; Murty, M.N. SSVM: A Simple SVM Algorithm. In Proceedings of the 2002 International Joint Conference on Neural Networks, Honolulu, HI, USA, 12–17 May 2002; Volume 3, pp. 2393–2398.
54. Mahesh, B. Machine learning algorithms—A review. *Int. J. Sci. Res. (IJSR)* **2020**, *9*, 381–386.
55. Orrù, P.F.; Zoccheddu, A.; Sassu, L.; Mattia, C.; Cozza, R.; Arena, S. Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry. *Sustainability* **2020**, *12*, 4776. [CrossRef]
56. Byvatov, E.; Fechner, U.; Sadowski, J.; Schneider, G. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882–1889. [CrossRef]
57. Evgeniou, T.; Pontil, M. Support vector machines: Theory and applications. In *Advanced Course on Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 249–257.
58. Zebari, R.Z.; Abdulazeez, A.M.; Zeebaree, D.Q.; Saeed, J.N.; Zebari, D.A. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 56–70. [CrossRef]
59. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the Science and Information Conference, London, UK, 27–29 August 2014; pp. 372–378.
60. Hira, Z.M.; Gillies, D.F. A review of feature selection and feature extraction methods applied on microarray data. *Adv. Bioinform.* **2015**, *2015*, 198363. [CrossRef] [PubMed]
61. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
62. Park, S.H.; Goo, J.M.; Jo, C.-H. Receiver operating characteristic (ROC) curve: Practical review for radiologists. *Korean J. Radiol.* **2004**, *5*, 11–18. [CrossRef] [PubMed]
63. Hoo, Z.H.; Candlish, J.; Teare, D. What is an ROC curve? *Emerg. Med. J.* **2017**, *34*, 357–359. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.