**DATA &
METADATA**

Check for
updates

ORIGINAL

Enhancing the hiring process: A predictive system for soft skills assessment

Mejorar el proceso de contratación: Un sistema predictivo para la evaluación de las competencias interpersonales

Asmaa Lamjid[1] ✉, Anass Ariss[1] ✉, Imane Ennejjai[1] ✉, Jamal Mabrouki[2] ✉, Soumia Ziti[1] ✉

[1]Department of Computer Science, Faculty of Sciences, Mohammed V University in Rabat, Rabat 10000, Morocco.
[2]Laboratory of Spectroscopy, Molecular Modelling, Materials, Nanomaterial, Water and Environment, CERNE2D, Mohammed V, University in Rabat, Faculty of Science, Rabat, Morocco.

ABSTRACT

Human Resource Management faces the ongoing challenge of identifying top-performing candidates to enhance organizational success. Traditional recruitment methods heavily rely on assessing hard skills alone, overlooking the importance of soft skills in identifying individuals who excel in their roles. To address this, our paper introduces a novel predictive model that leverages Artificial Intelligence in the hiring process. By analyzing soft skills extracted from CVs, cover letters, websites, professional social media, and psychometric tests, the model accurately predicts potential candidates suitable for specific job roles. This system effectively eliminates poor hiring decisions, reduces time and effort, minimizes recruitment costs, and mitigates turnover risks. The implementation of our proposed model employs various predictive machine learning classifiers, with key input soft skills including creativity, collaboration, empathy, curiosity, and critical thinking. Notably, the Support Vector Machine classifier emerges as the top-performing model in terms of predictive accuracy.

Keywords: Soft Skills; Recruitment Process; Machine Learning in Hiring; Artificial Intelligence; Predictive Hiring; Human Resource Management.

RESUMEN

La gestión de recursos humanos se enfrenta al reto constante de identificar a los candidatos más eficaces para mejorar el éxito de la organización. Los métodos tradicionales de selección de personal se basan en gran medida en la evaluación de las aptitudes físicas y pasan por alto la importancia de las aptitudes interpersonales a la hora de identificar a las personas que destacan en sus funciones. Para solucionar este problema, nuestro artículo presenta un novedoso modelo predictivo que aprovecha la Inteligencia Artificial en el proceso de contratación. Mediante el análisis de las competencias interpersonales extraídas de CV, cartas de presentación, sitios web, redes sociales profesionales y pruebas psicométricas, el modelo predice con exactitud los posibles candidatos adecuados para determinadas funciones. Este sistema elimina eficazmente las malas decisiones de contratación, reduce el tiempo y el esfuerzo, minimiza los costes de contratación y mitiga los riesgos de rotación. La aplicación de nuestro modelo propuesto emplea varios clasificadores de aprendizaje automático predictivo, con habilidades blandas de entrada clave como la creatividad, la colaboración, la empatía, la curiosidad y el pensamiento crítico. En particular, el clasificador Support Vector Machine resulta ser el modelo de mayor rendimiento en términos de precisión predictiva.

Palabras clave: Habilidades Interpersonales; Proceso de Contratación; Aprendizaje Automático en La Contratación; Inteligencia Artificial; Contratación Predictiva; Gestión de Recursos Humanos.

INTRODUCTION

In today's global business environment, organizations increasingly recognize the importance of human resources as a source of competitive advantage. The pressure to hire and retain high-performing, leadership-oriented employees has intensified. However, this task has become more challenging with many resumes on various platforms (applications, websites, and professional networks). Advancements in technology have revolutionized the hiring process, giving rise to predictive hiring—a method that utilizes artificial intelligence and data analysis to predict a candidate's potential success. Soft skills, such as creativity, collaboration, empathy, curiosity, and critical thinking, have become key indicators of a candidate's potential.

These skills are essential for succeeding in the future workplace and job advancement, as they increase productivity and effectiveness. This paper aims to explore the development of a new predictive hiring system based on soft skills, which can accurately and automatically predict potential candidates. The system encompasses every stage of predictive recruitment, from sourcing to predicting job performance. Soft skills, which can manifest in different linguistic forms depending on the context, pose unique challenges in this process.

By leveraging artificial intelligence and machine learning algorithms, vast amounts of data on a candidate's soft skills can be analyzed to provide valuable insights into their potential as an employee. This paper focuses on comparing three machine learning algorithms—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Trees—for their effectiveness in predictive hiring. The results demonstrate that the Support Vector Machine classifier outperformed the others in predicting a candidate's potential.

This innovative approach helps eliminate poor hiring decisions and reduces the time, effort, and costs associated with the traditional recruitment process while mitigating employee turnover. The paper provides a comprehensive overview of the predictive hiring process and its applications in the modern workforce. By examining the role of artificial intelligence and soft skills in predictive hiring, organizations can make more informed decisions, leading to better outcomes in talent acquisition.

Related works

Before starting our work, we studied the literature concerning our problematic general landscape. Indeed, we have first studied the scientific works in recruitment process automation, intelligent models, predictive systems based on artificial intelligence, and predictions of performances. However, most scientific research focused on predictive systems in the field of education by developing intelligent models for the prediction of the performance of students in different levels of study based on artificial intelligence is precisely machine learning. (Beth Dietz-Uhler & Janet E. Hurn, 2013) [1] discussed the significance of data analytics in predicting student performance. It emphasized students' interest levels, including factors like hobbies, ability, strengths, and leisure activities. (Roberto Bertolin & al., 2022) [2] discuss the variability in student performance predictions using nonparametric bootstrap algorithms in data pipelines. Bootstrapping was applied to examine performance variability among five data mining methods (DMMs) and four filter preprocessing feature selection techniques for forecasting course grades. The ensemble technique elastic net regression (GLMNET) significantly outperformed all other DMMs and exhibited the least variability in the AUC. However, all filter feature selection techniques significantly increased variability in student success predictions compared to when this step was omitted from the data pipeline. (Conijn et al., 2016) [3] developed a predictive model using a multilevel regression method with LMS data samples to gauge the job priorities of students. Various machine learning techniques are already used to predict a student's future. But most of those prediction models ignore the overall distribution differences between multiple clusters of students. (J. Kovacic, 2010) [4] discussed a real case scenario on educational information mining to detect the level of enrolment data usage on students' success rate. CART and CHAID algorithms were applied to classify successful and unsuccessful students. 59,4 % and 60,5 % was the accuracy rate obtained from CHAID and CART respectively. (Fadhilah, Aziz et al., 2015) [5] developed a computationally intelligent model to predict students' performance. WEKA software was used for the implementation. Attributes used for testing include family income level, college entry code, race, gender, regional place, and grade. The analysis showed that regional place, gender, and family income level were the most impactful attributes contributing to academic success.

Other works focus on developing an intelligent matchmaking model between the company's needs and the candidates applying for a specific offer. This matchmaking is based on the candidates' technical skills and other aspects such as age, gender, and years of experience to predict the candidate's performance and who will be the recruit in a given position. (Mustafa Agaoglu, 2016) [6] discussed several features and variables that were essential in determining employee performance. An attribute usage analysis was performed using many classifiers. The C5.0 algorithm showed a maximum usage of attributes compared to other classifiers like SVM and CART. (Guo et al., 2016) [7] undertook research and developed an automated model that mapped a candidate's resume to a potential job posting. Factors like job title, study area, competitive level, and education degree were accumulated from online job postings to develop this predictive job model. In a study

by (Cavnar and Trenkle, 2019) [8] n-grams approach was employed to classify job subjects. The dataset used constitutes 778 articles from 5 different newsgroups. The text data was available in seven class labels of the questionnaire. The classification accuracy could have been higher, ranging from 30 % to 80 %. (Sokolova et al., 2007) [9] analyzed the performance appraisal of employees by using a multi-class classification approach which selected a single class label among several labels. Micro averaging and macro-averaging methods were used to assess the performance of employees. Some analytical studies of high impact are also undertaken to determine factors influencing the career selection of college-going students. (Suryadi et al., 2020) [10] discussed the effect of a parent's career on choices made by students. According to (Dudley et al. 2006) [11] personality traits may be categorized into openness, conscientiousness, agreeableness, neuroticism, and extraversion. Among these traits, conscientiousness is the most crucial trait that determines the prospectiveness of a job. (Tongshan Chang & Ed.D, 2008) [12] introduced a real-time project module that assisted higher education institutions in achieving enrolment tasks using the machine learning approach. Moreover, the result confirmed that information mining has a vital role in the job recruitment process. (Ajay Kumar Pal & Saurabh Pal, 2013) [13] used classification algorithms like ID3, C4.5, and Bagging to determine students' performance during job interviews. They inferred that factors like qualification, grade, high school marks, and secondary level marks were highly correlated among them. (Sushruta Mishra et al., 2021) [14] developed an intelligent predictive model to decide upon a candidate's suitability for an applied It based job using the KNN (K-Nearest Neighbours) algorithm combined with a hard-voting approach is employed.

(Sridevi G.M & S. Kamala Suganthi, 2022) [15] developed an Artificial Intelligence (AI) system to measure and predict a suitable candidate from an available Candidate Resume (CR) database. The Jaccard similarity is measured between these clusters, and a suitability measure is proposed based on the cluster parameters. The prediction of candidate suitability is performed using the three classifiers: linear regression, decision tree, Adaboost, and XGBoost. Various features are formed by employing the bag of words technique to carry out the classification tasks. (Ayishathahira C H & al, 2018) [16] developed a system for resume parsing using deep learning models using the convolutional neural network (CNN), Bi-LSTM (Bidirectional Long Short-Term Memory), and Conditional Random Field (CRF) to classify a resume into three segments and extract 23 fields. (Marcu Florentina, 2020) [17] application the web scraping to extract a massive amount of data from websites using the UiPath automation tool.

The work conducted by (Ivo Wingsa & al, 2021) [18] presents a closely related approach to our proposed model. They propose a system that effectively extracts hard and soft skills from candidates' resumes and job descriptions using token classification. In another study, (Lamjid A and al., 2022) introduce a novel predictive micro-model specifically tailored for Information Technology consultants, utilizing soft skills as the basis for prediction. Furthermore, (Tongshan Chang & Ed.D24,2021) [20] introduce a real-time project module that employs machine learning to assist higher education institutions in enrollment tasks, emphasizing the crucial role of information mining in job recruitment. In other study, (Bodhvi Gaur and Al, 2021) [21] focuses on the challenge of extracting educational institutions' names and degrees from resume education sections. The authors propose a semi-supervised approach using a deep neural network model trained on a small annotated dataset. The model predicts entities in unlabeled sections, corrected by a module, and achieves 92,06 % accuracy through iterative training updates. Moreover, (Silvia Fareri and al, 2021) [22] introduced to automatically extract soft skills from text using a corpus of scientific papers.

Model description
Overview of the model

We have developed a new predictive hiring soft skills model that aims to forecast the performance of candidates for a particular job position. This model consists of six prototypes, as depicted in figure 1, which are as follows:

First, we commence with 'Sourcing & Data Collection' as we diligently gather diverse and comprehensive data from different sources. Subsequently, 'Data Analysis' takes center stage, where we delve deep into the acquired data to extract valuable insights, including variables, patterns, and trends that inform our hiring predictions. 'Data Pre-processing' emerges as the pivotal next step, involving data cleansing, transformation, and structuring to ensure its optimal readiness for input into our predictive model.

The heart of our predictive hiring system lies in 'Model Engineering & Execution,' where sophisticated models are meticulously crafted and executed, leveraging the refined data to predict candidate suitability. Lastly, 'Model Evaluation' brings our process to a close, subjecting our predictive model to rigorous scrutiny. We assess its accuracy, effectiveness, and performance in making informed hiring decisions, ensuring its reliability and practicality in real-world scenarios. This holistic approach ensures that our predictive hiring system is a robust and dependable tool for organizations seeking to build effective, well-rounded teams.
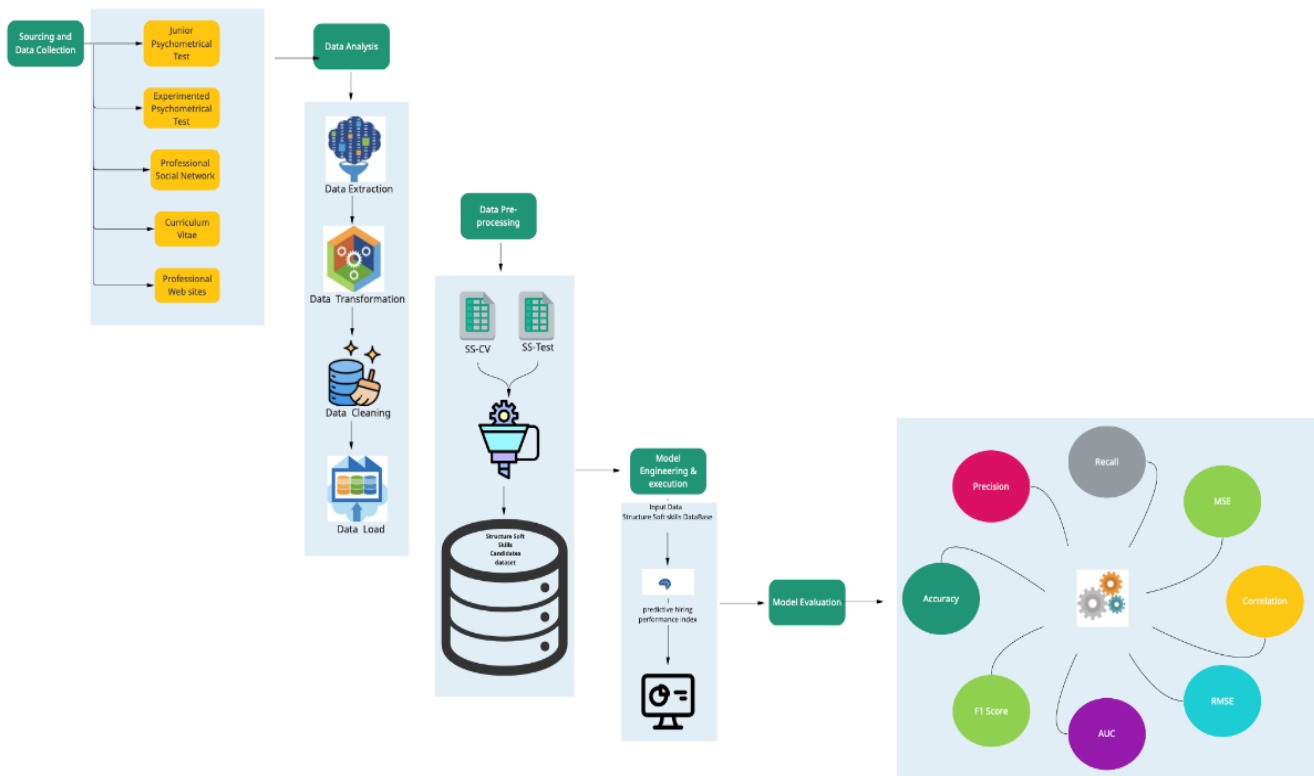
Figure 1. New Predictive Hiring Soft Skills Model
Source: This figure is self-designed and produced to illustrate our research concepts and results.

Delving Deeper: A Comprehensive Exploration of Model Details
Sourcing and Data collection
The sourcing and data collection phase holds significant importance as it caters to the unique requirements of each company's recruitment process. Every company seeks distinct profiles and specialties while adhering to selection criteria based on skills, abilities, and motivations. When a recruiter is faced with a vast pool of candidates, the predictive recruitment model becomes invaluable. It enables the selection of a limited number of candidates possessing all the necessary qualifications and soft skills essential for the job.

Data Analysis
The second phase focuses on three essential phases: data extraction, transformation, and loading.

Data extraction
Extracting soft skills from various sources poses significant challenges for us, primarily due to their diverse linguistic forms depending on the context. Hence, this step focuses on extracting soft skills from resumes, professional networks, and psychometric tests, employing three different methods. The data extraction process is outlined as follows:
1. Extracting data from resumes: We utilize a technique called resume parsing to extract soft skills from candidate resumes.
2. Extracting data from professional networks: We employ data scraping techniques to gather data from websites and store it in the desired format.
3. Extracting data from psychometric tests: We employ intelligent question answering to determine the ranking of soft skills for each candidate.

It is important to note that the psychometric test plays a crucial role in extracting candidates' soft skills by placing them in real professional scenarios. Consequently, the test's development was a complex phase completed in collaboration with psychiatrists and coaching specialists who specialize in the human aspects.

The data collected from the aforementioned process comprises information on the candidates' soft skills, such as collaboration, creativity, empathy, critical thinking, and curiosity. These five soft skills are vital for every employee and candidate to possess, develop, and demonstrate for success and career advancement.

Data Transformation
During this phase, we convert professional social network profiles and resumes, which are typically

encoded in file formats such as PDF, DOCX, ODT, DOC, and RTF, into a text format file.

Data Cleaning

Following the data transformation phase, the resulting file may contain unwanted lines, punctuations, bullets, and other irrelevant elements. We employ string replacement methods and regular expressions to eliminate these elements from the data.

Data Load

Data loading involves transferring the data sets obtained from the previous phases into the soft skills databases. Resumes and professional network data will be stored in a Cv-soft skills database. On the other hand, soft skills identified through the psychometric test will be stored in a database named Test-soft skills. It is worth mentioning that soft skills derived from the psychometric test will be converted into weights ranging from 1 to 5

Data Pre-processing

During this phase, we compare the soft skills stored in the Cv-soft skills and Test-soft skills databases with those extracted from the psychometric test. If the soft skills obtained from the CVs and professional social media profiles match the soft skills collected from the psychometric test, we calculate a performance index ranging from 1 to 4. A performance index of 1 represents a low-performance level, 2 indicates a medium-performance level, 3 represents a high-performance level, and 4 indicates an exceptional performance level.

However, if the soft skills do not match, we prioritize the soft skills extracted from the psychometric test. The purpose of this step is to identify the relevant soft skills of job seekers by assessing them in realistic professional scenarios.

After this filtering process, we obtain a structured database called the structured soft skills database, which we will utilize in the subsequent step.

Model Engineering and execution

In this phase, we deploy a combination of three machine learning algorithms: Decision Tree, Support Vector Machine, and K-Nearest Neighbors. The first step is to determine an appropriate split between training and test data, considering the available dataset. In our case, we allocate 80 % of the data for training and 20 % for testing purposes. Next, we preprocess the training and test data, ensuring that the model is trained using the structured soft skills data. The trained model is then used to make predictions for potential candidates. It's important to note that the model can dynamically receive input data (soft skills) and update the prediction results automatically, allowing for continuous improvement and adaptability

Model Evaluation

The evaluation of the models is conducted based on their performance and efficiency. In our case, we assessed the model using twelve evaluation metrics, ranging from accuracy to correlation. These metrics provide insights into the model's performance in terms of its ability to make accurate predictions and its correlation with the actual outcomes. By considering multiple evaluation metrics, we gain a comprehensive understanding of the model's effectiveness and suitability for the given task.

Presentation of the results

This section provides a description of the algorithms employed for predicting the performance index, including SVM, decision tree, and KNN. It also includes a discussion on the comparison of the results obtained.

Svm soft skills predict hiring implementation
SVM implementation before optimization

The SVM algorithm provides the flexibility to select various kernel functions for its processing. Kernel functions are responsible for mapping the data into higher dimensional spaces, a process commonly referred to as "kernelling." Different types of kernel functions exist, including Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. Each function possesses its own unique characteristics, advantages, disadvantages, and corresponding equation. However, determining the most effective function can be challenging. To address this, it is customary to choose different kernel functions and compare their results using a given dataset. In this case, we will utilize the default function, RBF (Radial Basis Function). Once the accuracy of the four functions has been compared, the results can be assessed.

Table 1. SVM results before optimization				
Metrics	**SVM**			
	RBF	**LIN**	**POLY**	**HINGE**
Accuracy - Train	0,80	0,79	0,70	0,75
Accuracy - Test	0,87	0,70	1,0	0,76
Recall	0,84	0,85	0,76	0,56
Precision	0,77	0,73	0,63	0,54
MSE - Train	0,09	0,125	0,0	0,23
MSE - Test	0,20	0,25	0,60	0,60
MAE - Train	0,20	0,25	0,40	0,25
MAE - Test	0,20	0,25	0,40	0,25
F1-Score	0,81	0,79	0,72	0,73
Correlation	0,72	0,70	0,60	0,60
Source: This table is self-designed and produced to illustrate our research concepts and results.				

Although the SVM model with the RBF function produces the best outcome, attaining an accuracy of 80 % is deemed inadequate. Hence, our goal is to improve the algorithm's performance to exceed a 90 % accuracy threshold. To accomplish this, we employ optimization techniques focused on minimizing the log loss function. The primary aim of these techniques is to enhance the model's overall accuracy and minimize the occurrence of false predictions.

SVM implementation after optimization
 a) Training SVM-RBF
In this phase, we begin by initializing the SVM classifier object named 'clf'. The chosen kernel for this classifier is 'RDF'. Additionally, we specify several parameters, including a gamma value of 0,3, a regularization parameter 'C', a maximum iteration count of 20, and set the 'probability' parameter to True.

After the initialization, we proceed to train the 'clf' object using the 'fit' method. The training data, represented by 'XSVM_train', along with its corresponding labels, denoted as 'YSVM_train', are provided as inputs to this method. By doing so, the classifier learns from the provided training data and labels, acquiring the ability to make predictions based on the learned patterns and relationships within the data.

 b) Prediction SVM-RBF example

Table 2. Illustration of the first 5 rows of actual and predicted data from the structured "soft skills database"	
Actual performance Index	**Predicted performance Index**
1	1
3	3
1	1
1	1
1	1
Source: This table is self-designed and produced to illustrate our research concepts and results.	

By comparing the model's predictions on the testing set with the actual set, we verified its robustness and reliability in making accurate assessments.

Evaluation of the SVM-RBF model

Table 3. results of evaluation metrics of SVM-RBF	
Evaluation Metrics	**SVM-RBF**
Accuracy training	0,95
Accuracy testing	1,0
Precision	1,0
Recall	1,0
F1 score	1,0
AUC Score	1,0
Correlation	1,0
MSE training	0,0

MSE testing	0,0
MAE training	0,0
MAE testing	0,0
RMSE training	0,0
RMSE testing	0,0
Source: This table is self-designed and produced to illustrate our research concepts and results.	

The SVM model was evaluated for its training and testing accuracies, which yielded a training accuracy of 0,95 and a testing accuracy of 1,0. These accuracy assessments demonstrate the model's ability to effectively classify instances in both the training and testing datasets. A training accuracy of 0,95 indicates excellent performance in accurately predicting labels for the training data. Furthermore, the model exhibited impressive performance, with minimal values of MSE, MAE, and RMSE between the training and testing data. Additionally, it achieved low recall, precision, and AUC scores using the one-vs-one (ovo) strategy. These outcomes highlight the model's effectiveness in correctly classifying instances in unseen testing data while successfully addressing the issue of overfitting. The high accuracy scores underscore the model's reliability, robustness, and potential for practical applications in real-world scenarios.
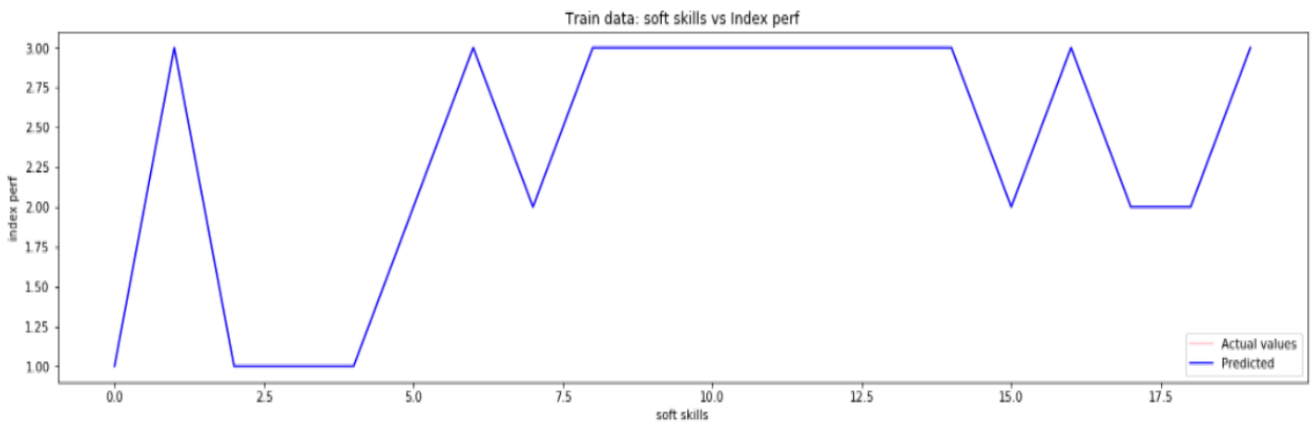

Figure 2. Plot performance index soft skills predicted vs. actual
Source: This figure is self-designed and produced to illustrate our research concepts and results.

KNN soft skills predict hiring implementation
Evaluation of the KNN model
    K-Nearest Neighbors (KNN) algorithm uses similarity calculations to predict based on input and training data. Randomly setting K as 2 resulted in unsatisfactory evaluation metrics, indicating potential overfitting due to the disparity between training and testing outcomes. The figure 3 below visualizes the predicted and actual data of the model, with a training accuracy of 0,90 and a test accuracy of 0,85.
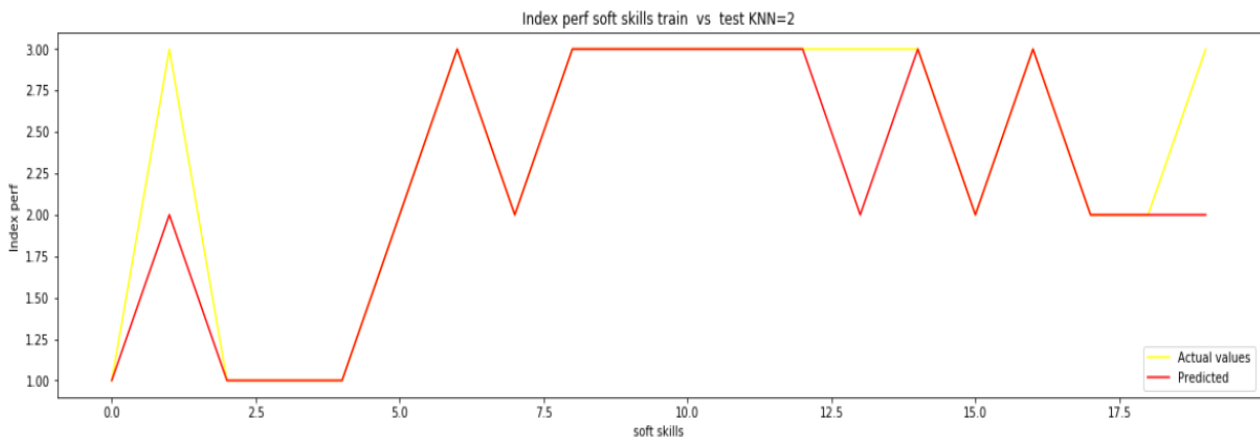

Figure 3. Plot the KNN soft skills predicted vs the KNN soft skills actual
Source: This figure is self-designed and produced to illustrate our research concepts and results.

Evaluation of the KNN model with K=1

To enhance the model, we developed a method to automatically predict the optimal value of K based on desired evaluation metrics for the KNN predictive soft skills model. After multiple iterations, our method determined K=1, as depicted in figure 4.
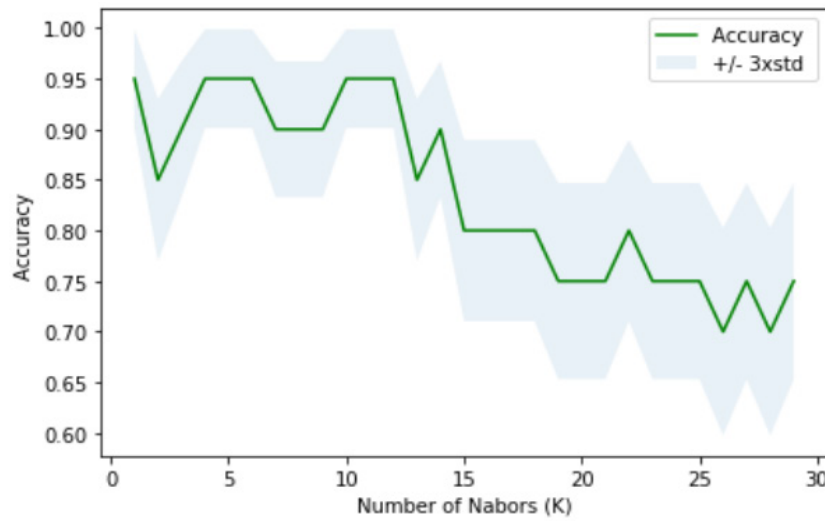

Figure 4. Determining the Optimal K for Maximum Accuracy in KNN Calculation Outputs
Source: This figure is self-designed and produced to illustrate our research concepts and results.

To assess the robustness of the KNN predictive hiring soft skills model, we examined it using the following 13 evaluation metrics:

Table 4. Results of evaluation metrics of KNN soft skills model	
Evaluation Metrics	**KNN-K1**
Accuracy training	1,0
Accuracy testing	0,95
Precision	0,97
Recall	0,97
F1 score	0,97
AUC Score	0,99
Correlation	0,92
MSE training	0,0
MSE testing	0,05
MAE training	0,05
MAE testing	0,05
RMSE training	0,0
RMSE testing	0,02
Source: This table is self-designed and produced to illustrate our research concepts and results.	

The results obtained for KNN with K equal to 1 showcase the model's exceptional predictive capabilities. The accuracy achieved by the model indicates its ability to make highly accurate predictions, which has significant implications for streamlining and optimizing the recruitment process. By accurately identifying and selecting the right individuals for the appropriate positions, the model minimizes the time and effort spent on recruitment, ensuring a more efficient and effective hiring process. This not only saves valuable resources but also enhances the overall quality of the workforce by matching candidates with the most suitable roles. Ultimately, the accurate predictions with 1,0 for training, 0,95 for testing and the minimal error rate between the training and testing datasets provided by the KNN-K1 model contribute to improved organizational performance and productivity.

Decision tree soft skills predict hiring implementation

Decision trees are nonparametric supervised learning methods that use a tree-like structure for decision-making and prediction. They offer interpretable decision rules and are constructed in a top-down manner. Decision trees are valuable tools for understanding data logic and making predictions in diverse fields. In our case, we initially deployed the Decision Tree soft skills model using default parameters, specifically setting the random state to 3 and entropy to 2. However, upon evaluating the model's performance, we observed a training accuracy of 0,76 and a testing accuracy of 0,5. These relatively low accuracy scores indicate that the model does not accurately predict candidates' soft skills.

The figure 5 below visualizes the predicted and actual data of the model, with a training accuracy of 0,76 and a test accuracy of 0,50.

Figure 5. Plot the Decision Tree performance index soft skills predicted vs actual before optimization
Source: This figure is self-designed and produced to illustrate our research concepts and results.

Evaluation of the decision tree predictive soft skills model

a) Enhanced Performance of Decision Tree with Optimized Parameters: Random State = 13 and Max-Depth Entropy = 4

In pursuit of performance improvement, we optimized the model's parameters by carefully determining the optimal random state (13) and maximum depth entropy (4) through rigorous experimentation. These optimized settings were then applied to re-deploy the Decision Tree soft skills model. The subsequent table presents the enhanced performance and predictive accuracy achieved as a result of this implementation.

Table 5. Results of evaluation metrics of KNN soft skills model	
Evaluation Metrics	**Decision Tree**
Accuracy training	0,90
Accuracy testing	0,95
Precision	0,93
Recall	0,93
F1 score	0,93
AUC Score	0,96
Correlation	0,93
MSE training	0,01
MSE testing	0,05
MAE training	0,05
MAE testing	0,05
RMSE training	0,032
RMSE testing	0,025
Source: This table is self-designed and produced to illustrate our research concepts and results.	

The figure 6 below visualizes the predicted and actual data of the model after augmentation of the parameters.

Figure 6. Plot the Decision Tree performance index soft skills predicted vs actual before optimization
Source: This figure is self-designed and produced to illustrate our research concepts and results.

Based on our findings, we can confidently conclude that the optimized model demonstrates exceptional performance in predicting candidates. With an impressive accuracy rate of 0,95 and 0,90 for training and testing, respectively, the model consistently achieves highly accurate predictions. Additionally, the minimal error rates observed between the predicted and actual data further confirm the model's reliability and precision.

These results highlight the effectiveness of the optimization process in improving the model's predictive capabilities. By fine-tuning the parameters and refining the algorithm, we have significantly enhanced the model's accuracy and minimized potential errors. As a result, the model now provides reliable and trustworthy predictions, enabling more informed decision-making in candidate selection.

The high accuracy scores and minimal error rates validate the model's robustness and effectiveness, making it a valuable tool in the candidate evaluation process. Organizations can rely on this optimized model to make accurate predictions, streamline their hiring processes, and ultimately secure the most suitable candidates for various positions.

DISCUSSIONS AND RESULT

The table below represents a resume of the 3 implementations of the soft skills models.

Table 6. Summary of the predictive hiring soft skills models						
P-Metrics			**SSML-Models**			
	Decision Tree		**K- Nearest Neighbors**		**Support Vector Machine**	
	BOP	**AOP**	**BOP**	**AOP**	**BOP-RBF**	**AOP**
Accuracy-Test	0,5	0,95	0,85	0,95	0,80	1
Accuracy-Train	0,76	0,9	0,9	1	0,87	0,95
Recall	0,33	0,93	0,87	0,97	0,84	1
Precision	0,33	0,93	0,87	0,97	0,77	1
MSE-Test	0,65	0,05	0,15	0,05	0,09	0
MSE-Train	0,237	0,01	0,1	0	0,20	0
MAE-Test	0,55	0,05	0,15	0,05	0,20	0
MAE-Train	0,55	0,05	0,15	0,05	0,20	0
RMSE-Test	0,8	0,022	0,39	0,02	-	0
RMSE-Train	0,487	0,031	0,31	0	-	0
F1-Score	0,34	0,93	0,87	0,97	0,81	1
AUC-Score	0,68	0,96	0,81	0,99	-	1
Correlation	0,17	0,93	0,79	0,92	0,72	1
Source: This table is self-designed and produced to illustrate our research concepts and results.						

- **BOP = BEFORE OPTIMISATION**
- **AOP = AFTER OPTIMISATION**

The success of a company relies on the quality of its hires, their training, and retention. With digitalization, HR practices have evolved, necessitating the adoption of digital recruitment, career development, and

streamlined employee-organization interactions. This has resulted in an abundance of resume data from various platforms. Traditional hiring processes can be inefficient and costly when dealing with a large number of candidates. To address these challenges, a new predictive hiring model focusing on soft skills, alongside hard skills, was developed. It streamlines the recruitment process, reduces time-to-hire, turnover rates, and costs, enhances hiring decisions, and predicts future employee performance. The model collects and structures data from various sources, conducts pre-selection based on company requirements, and evaluates candidates' soft skills through psychometric tests. Soft skills data undergoes preprocessing to ensure accuracy and reliability.

The model incorporates three machine learning algorithms: decision tree, K-nearest neighbors (KNN), and support vector machines (SVM) with different kernel functions. SVM with the radial basis function (SVM-RBF) achieved the highest accuracy at 80 %. Further performance improvements were pursued across all three models to reduce overfitting and minimize errors between training and testing datasets. Evaluation metrics including accuracy, precision, recall, F1-Score, RMSE, MAE, MSE, AUC Score, and correlation were used to assess model performance. A dataset of 2000 resumes was split into 80 % for training and 20 % for testing. The SVM-RBF model demonstrated a perfect testing accuracy of 100 %, while both KNN and Decision Tree models achieved a testing accuracy of 95 %. Based on these results, the SVM-RBF model proved to be the most reliable for predicting the performance index of selected candidates.

CONCLUSIONS

Our predictive hiring soft skills model revolutionizes candidate evaluation by recognizing the crucial role of soft skills in today's work environments. Our model provides invaluable data-driven insights by integrating cutting-edge prototypes and leveraging sophisticated algorithms.

Through accurate predictions and comprehensive assessments of essential soft skills, our model enables organizations to build effective and well-rounded teams. It ensures that candidates possess the technical expertise and interpersonal abilities to thrive. With streamlined recruitment processes and a focus on soft skills, our model empowers organizations to identify and evaluate qualified candidates confidently. This fosters collaboration, adaptability, critical mind, empathy, and innovation in highly skilled workforces.

In summary, our predictive hiring soft skills model offers a game-changing solution, emphasizing the critical role of soft skills. Organizations make informed hiring decisions by leveraging data-driven insights, leading to enhanced performance and long-term success.

Our model achieves exceptional accuracy, enables confident candidate selection, and optimizes recruitment. It aligns talent acquisition with organizational objectives, driving success and achieving long-term goals.

In the upcoming phases of our research, our primary focus will be on fully implementing the entire model while harnessing the power of a diverse combination of more than three algorithms. This strategic approach aims to enhance the model's predictive capabilities and further optimize its performance, ensuring it remains at the forefront of cutting-edge technology in predictive hiring systems.

REFERENCES

1. Beth Dietz-Uhler & Janet E. Hurn, « Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective », Journal of Interactive Online Learning. Volume 12, Number 1, 2013 ISSN : 1541-4914 ;

2. Df Roberto Bertolini, Stephen J. Finch & Ross H. Nehm, « Quantifying variability in predictions of student performance : Examining the impact of bootstrap resampling in data pipelines », Computers and Education : Artificial Intelligence Volume 3, 2022, 100067 ;

3. Rianne Conijn, Ad Kleingeld & Uwe Matzat, « Predicting Student Performance from LMS Data : A Comparison of 17 Blended Courses Using Moodle LMS », October 2016, IEEE Transactions on Learning Technologies PP (99) : 1-1 ;

4. Zlatko J. Kovacic, « Early Prediction of Student Success : Mining Students Enrolment Data ». January 2010, DOI : 10.28945/1281 Conference : InSITE 2010 : Informing Science + IT Education Conference ;

5. ARISS, Anass, ENNEJJAI, Imane, MABROUKI, Jamal, et al. Tracking System for Living Beings and Objects: Integration of Accessible Mathematical Contributions and Graph Theory in Tracking System Design. Data and Metadata, 2024, vol. 3, p. . 376-. 376;

6. Mustafa Agaoglu, « Predicting Instructor Performance Using Data Mining Techniques in Higher Education », Department of Computer Engineering, Marmara University, anuary 2016 IEEE Access 4 :1-1 ;

7. Shiqiang Guo, Folami Alamudun & Tracy Hammond, « RésuMatcher : A personalized résumé-job matching

system ». Expert Systems with Applications Volume 60, 30 October 2016, Pages 169-182 ;

8. William B. Cavnar and John M. Trenkle, « N-Gram-Based Text Categorization ». 2019, Environmental Research Institute of Michigan P.O. Box 134001 Ann Arbor MI 48113-4001 ;

9. MABROUKI, Jamal, BENBOUZID, Maria, DHIBA, Driss, et al. Internet of things for monitoring and detection of agricultural production. In : Intelligent Systems in Big Data, Semantic Web and Machine Learning. Cham : Springer International Publishing, 2021. p. 271-282;

10. Bambang Suryadi, Bahrul Hayat & Muhammad Dwirifqi Kharisma Putra, « The Influence of Adolescent-Parent Career Congruence and Counselor Roles in Vocational Guidance on the Career Orientation of Students ». April 2020 International Journal of Instruction 13 (2) : 45-60 ;

11. Dudley, Nicole M. Orvis, Karin A. Lebiecki, Justin E. Cortina & José M. « A meta-analytic investigation of conscientiousness in the prediction of job performance : Examining the intercorrelations and the incremental validity of narrow traits ». 2006, Journal of Applied Psychology, 91(1) ;

12. ENNEJJAI, Imane, ARISS, Anass, MABROUKI, Jamal, et al. Enhancing Misinformation Detection Using Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) with Word Embedding Techniques. Discrete Mathematics, Algorithms and Applications, 2024 ;

13. ENNEJJAI, Imane, ARISS, Anass, MABROUKI, Jamal, et al. An Artificial intelligence Approach to Fake News Detection in the Context of the Morocco Earthquake. Data and Metadata, 2024, vol. 3, p. . 377- 377;

14. Sushruta Mishra, Pradeep K Mallick, Hrudaya K Tripathy, Lambodar Jena, and Gyoo-Soo Chae. « Stacked KNN with hard voting predictive approach to assist hiring process in IT organizations ». International Journal of Electrical Engineering & Education. 2021 ;

15. Sridevi G, M & S. Kamala Suganthi. « AI based suitability measurement and prediction between job description and job seeker profiles ». International Journal of Information Management Data Insights Volume 2, Issue 2, November 2022, 100109 ;

16. C H Ayishathahira, C Sreejith & C Raseek. « Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing ». 2018 International CET Conference on Control, Communication, and Computing (IC4). INSPEC Accession Number : 18233630 ;

17. LAMJID, Asmaa, ARISS, Anass, ENNEJJAI, Imane, et al. Predictive Hiring Micro Systems for Data Analysts Through Soft Skills Assessment. In : Technical and Technological Solutions Towards a Sustainable Society and Circular Economy. Cham : Springer Nature Switzerland, 2024. p. 295-302;

18. Ivo Wings, Rohan Nanda & Kolawole John Adebayo. « A Context-Aware Approach for Extracting Hard and Soft Skills ». Procedia Computer Science Volume 193, 2021, Pages 163-172 ;

19. Asmaa Lamjid, Karim El Bouchti, Soumia Ziti, Reda Oussama Mohamed, Hicham Labrim, Anouar Riadsolh & Mourad Belkacemi. « Predictive Hiring System : Information Technology Consultants Soft Skills », 2022 : International Conference on Advanced Intelligent Systems for Sustainable Development pp 680–685

20. AZROUR, Mourade, MABROUKI, Jamal, FARHAOUI, Yousef, et al. Security analysis of Nikooghadam et al.'s authentication protocol for cloud-IoT. In : Intelligent systems in big data, semantic web and machine learning. Cham : Springer International Publishing, 2021. p. 261-269;

21. Bodhvi Gaur, Gurpreet Singh Saluja, Hamsa Bharathi Sivakumar and Sanjay Singh. « Semi-supervised deep learning based named entity recognition model to parse education section of resumes ». Neural Computing and Applications volume 33.2021, 5705–5718 ;

22. Silvia Fareri, Nicola Melluso, Filippo Chiarello and Gualtiero Fantoni. « SkillNER: Mining and mapping soft skills from any text ». Expert Systems with Applications Volume 184,2021.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Asmaa Lamjid, Anass Ariss, Imane Ennejjai.
Data curation: Asmaa Lamjid, Anass Ariss, Imane Ennejjai, Jamal Mabrouki, Soumia Ziti.
Formal analysis: Asmaa Lamjid, Anass Ariss, Imane Ennejjai, Soumia Ziti.
Research: Asmaa Lamjid, Anass Ariss, Imane Ennejjai, Jamal Mabrouki, Soumia Ziti.
Methodology: Asmaa Lamjid, Anass Ariss, Imane Ennejjai, Jamal Mabrouki, Soumia Ziti.
Supervision: Soumia Ziti.
Validation: Asmaa Lamjid, Anass Ariss, Soumia Ziti.
Drafting - original draft: Asmaa Lamjid, Anass Ariss, Imane Ennejjai.
Writing - proofreading and editing: Asmaa Lamjid, Anass Ariss, Jamal Mabrouki, Soumia Ziti.