Use of data mining to optimize the recruitment process in a software development company

K.Aradika¹, G.Geethanjali², R.Kiruthika³, J.Priyadharshini⁴, J.R.Thresphine⁵ ^{1,2,3,4,5}Department of Computer Science and Engineering, Prist Deemed To Be University, Vallam, Thanjavur

Abstract

Data mining is a branch of artificial intelligence that has many current applications. It is mainly used for reliable predictions of events based on historical data. This article presents a proposal to optimize the personnel selection processes in a software development company with the aim of accelerating the evaluation process that will help them to quickly identify the candidates with the greatest potential.

Keywords: Artificial Intelligence, Data Mining, Recruitment.

1. Introduction

The human capital, referring to the investment in education and training of the personnel of a company that increases the productivity of each individual [1], is one of the basic competences for technology companies to maintain their competitive advantage in an area where regulates the acquisition and updating of knowledge. The hiring of qualified personnel is affected by the recruitment process of the companies, since the necessarily exhaustive evaluations require a sequence of evaluations that make the procedures last for several weeks.

The work posts require personnel with more and diverse experience, which is why the selection process is more complex. Therefore, the conventional approaches to staff selection that are developed on the basis of the characteristics of static work will no longer be sufficient [2]. In order to find the people indicated for a post, it is vital to develop effective selection approaches. The use of data mining in this area seeks to find employers that distinguish the best candidates so that the selection processes are dynamically adjusted to direct the efforts to those professionals with the greatest probability of being successful workers in the organization.

This document presents the theoretical framework in section 2, the background and the description of the problem in section 3. Section 4 describes the proposal to implement for the solution of the problem and section 5 covers the expected results and benefits. Finally, we close with a series of conclusions.

2. Theoretical framework

The selection of personnel plays a decisive role in the management of human resources to determine the quality of the hired personnel. Previous investigations [3-4] have reviewed staff selection studies and found that situations such as change in organizations, roles and work standards have influenced hiring processes. While advances in information technologies also affect candidate selection in human resource management [5-6]. The application of expert systems to help in decision-making in recruitment has grown [8-9] and organizations are exploring new options to improve their staff selection processes and direct their efforts towards the optimal candidates for each position.

Due to advances in information technologies, researchers have developed systems to help in decision-making and expert systems to improve the results of human resource management. In particular, data mining is recognized as one of the most outstanding disciplines, as it refers to the extraction or use of patterns from a

large data base through automatic or semi-automatic exploration and analysis [10-11] These methodologies have been developed for the exploration and analysis of large amounts of data to discover significant patrons and rules. The data, including personal information, can be a valuable resource for knowledge discovery and decision-making support. In a study by the National Tsing Hua University, a strategy was developed to extract useful rules from the relationships between personal profile data and their behaviour at work by means of decision trees [12]. The results set out rules to be followed during the selection procedure to ensure timely follow-up of the best candidates. Our focus will try to evaluate different algorithms to identify the hidden patrons behind the results of the interviews and selection exams.

3. Problematic

The investigation will take place in a Mexican outsourcing company that is dedicated to software development and provides its services to international companies where the evaluation standards are high, so that the quality of its employees is directly reflected in the results. The selection of candidates is crucial to achieve the objectives, maintain the relationship with the client and take care of the prestige of the company for the growth of the business. The recruitment process is complex, as it consists of 7 stages through which each candidate must pass: First contact (1), logic exam (2), interview in English (3), light technical interview (4), in-depth technical interview (5), cultural interview (6) and joint programming (7). Each of the recruitment stages qualifies on the basis of a numerical scale and each of the qualifications is accompanied by a brief justification of why the people who were involved in that part of the interview assigned this value; regularly is an average of 6 people different including part of the team to which the candidate aspires and the recruiter. Each of the stages of the recruitment process requires an investment of resources by both parties. The average time reversed to a candidate is 8 hours; they are from the first contact until they accept the offer. All this time extends to 5 weeks on average.

4. Solution proposal

Based on the above and given that the company does not have a way of knowing which of the candidat es will pass to the final stage, the human resources department cannot prioritize its recruitment processes and this has caused that many times it loses applicants with more experience and better preparation, because during their application process they are contacted by competent companies. Our proposal consists of the use of a supervised artificial intelligence learning algorithm capable of detecting behaviour patterns in the evaluations of each one of the interviewees in early stages and predicting which of them have a greater probability of being a failure. successful lador within the company, in order to be able to prioritize your recruitment process. The company has a performance history in evaluations during the selection process for each candidate. Some of the candidates were discarded in saying Ferentes stages, siendo those who will arrive at the end, are able to occupy a place.

The history of results of the exams applied to the candidates will be used to train the models. Among the data that form part of the representation of the instances are found the score of each evaluation and in a more granular way the answers for each one of the reagents of the first of the exams presented.

Candidates will be classified in 5 classes, where each class indicates the level each one will reach during the process. Two classification models will be evaluated: the vector support machine model and a naive Bayes classifier. Both results will be compared with the performance of decision trees. The models and the training and evaluation procedure are described below.

The personal data and demographic characteristics of the individuals will not be taken into account during this stage, as our results will be based exclusively on their ability to solve problems.

4.1 Decision trees

One of the most used algorithms for decision-making at a business level is the decision tree. Its function is to evaluate the groups of data and produce rules that help classify a new class (Figure 1a). In our proposal, the vector support machine models and a naive Bayes classifier will be compared with the classification results by means of decision trees.



Fig. 1. a) Traditional application: decision trees.b) Model training using a vector support machine (SVM).





4.2 The Vectorial Support Machine Model

One of the proposed models for the early identification of the optimal candidates for the proposal consists of a vector support machine (SVM by its abbreviations in English). The multiclass model (5 classes) will be trained with the exam history of the different candidates (Figure 1b) to be tested with the classification of new instances, this is, the prediction of the level at which a candidate can get based on the results of from the first exams.

4.3 Naive Bayes Classifier

In addition to the previous models, a naive Bayes classifier will be trained with the histories to evaluate new instances and predict the probability that a candidate reaches a certain level. Despite its simplicity, the classifier has demonstrated its potency in the classification of instances assuming the independence between the events, which in this case are the answers to the questions of the first of the exams.

Once the candidate presents his first evaluations his results are evaluated by the previously trained model. The result will be a vector with the probability of belonging of the candidate to the class according to his/her performance (Figure 2). This result will allow applying criteria for the selection of the candidate.

5. Expected results and benefits

The research project will evaluate several options to select an adequate algorithm for the historical data with which the development company counts on its candidates with the aim of being implemented and that future aspirants will be submitted to be able to predict their success in the process. Our proposal intends to overcome the results of a decision tree, since it needs instances that cover all possibilities in the resolution of an exam, and many times the data does not exist. On the other hand, the implementation of a machine model with vectorial support or a naive Bayes model will allow the generation of hierarchical models in the future to improve the performance of the classifier. The benefits of this implementation are translated into a timely detection of the most competent candidates to forward efforts to their hiring and avoid the risk of losing them.

6. Conclusion

The recruitment team and the selection of personnel play a role that is directly reflected in the quality and performance of the workers. These processes are currently reinforced by techniques such as decision-making help and information technologies. A classification model for the detection of the best candidates in their first evaluations could revolutionize recruitment procedures, making them more dynamic and efficient by evaluating the hiring history and current information on candidates. The mining of data can play an important role in helping the selection processes to predict which candidates follow a pattern of success and in this way identify these key elements to give them priority over others, reducing the risk of losing them during the process, improving it quality of the employees that will result in better products in the company, an increase in earnings and reduction of costs in the long term.

References

- 1) Goldin, C.: Human capital. Handbook of Cliometrics, 55-86 (2016).
- Borman, W.C., Hanson, M.A., Hedge, J.W.: Personnel selection. Annual review of psychology 48, 299–337 (1997).
- Robertson, I.T., Smith, M.: Personnel selection. Journal of occupational and organizational psychology 74, 441–472 (2001).
- Beckers, A. M., Bsat, M. Z.: A DSS classification model for research in human resource information systems. Information Systems Management 19, 41–50 (2002).
- Valcárcel Asentios, V. (December 2004). Data Mining and the Discovery of Knowledge. Revista de la Facultad de Ingeniería Industrial UNMSM, 4.
- 6) M S Chen, Han J W, Yu P S. Data Mining: An Overview from a Database Perspective. . IEEE Transaction on Knowledge and Data Engineering. 1996, 18 (6): 1-41.

- 7) Y.B.Qu,H.Zheng. Database Marketing Beijing: Management Press, 2009.
- P.Zheng. Data Mining and Its Application in Marketing . The Journal of Beijing Institute of Light Industry, 2010, 3
- 9) W.W.Chen, W. The Course of Data Warehouse and Data Mining. BeijingTsinghua Press2006.
- Eckhorn R,Reitboeck H J, Arndtm,et al. A neural network for feature linking via synchronous activity: results from cat visual cortex and from simula-tions[M] Cambridge: Cambridge University Press,1989, 255-272.
- 11) Chalotra, P., & Dutta, M. (December 2011). researchgate.net. Retrieved on February 18, 2016, from https://www.researchgate.net/publication/216700550_Performance_Analysis_of_ Clustering_Methods_for_Outlier_Detection
- 12) Cravero Leal, A., & Sepúlveda Cuevas, S. (2009). Application of Data Mining for the Detection of Anomalies: A Study Case. International Workshop, 8.