

Job Recruitment Analysis based on Xgboost Decision Tree

Jiawen Chen, Yiqi Chen, Yunyang Liao, Juanjuan Mu, Guowei Li, Juan Li*

Wuhan Business University, Wuhan, China

*Corresponding Author

ABSTRACT

With the increasing number of college students, "the most difficult graduation season" is a topic of anxiety for many parents and children. Therefore, before looking for a job, Xgboost decision tree analysis of the position is of great significance to job seekers and work units. In view of the diversity of region, position, salary and education, Xgboost decision tree algorithm can visually show information to job seekers in the way of word cloud for job seekers, so as to facilitate the corresponding job recommendation for job seekers.

KEYWORDS

Xgboost Decision Tree; Job Application Analysis; Visualization; Word Cloud.

1. INTRODUCTION

According to the latest data released by the Ministry of Education, the number of college graduates in 2022 reached a new high of more than 10 million, and the employment pressure of the whole society is still huge[1]. To solve these problems, it is necessary to collect and analyze the data of various aspects such as the demand and some requirements of the recruitment positions, so as to provide employment guidance for college graduates. Therefore, it is very important to carry out certain analysis in the job application. It cannot only make job seekers more intuitive to understand the current employment market and demand, but also make job seekers more quickly and convenient to obtain suitable for their own job demand information.

In the job application analysis work, adhere to the principles of meticulousness and science, identify and identify the function setting, and take the job analysis results as the guiding basis, which is conducive to promoting the smooth[2] development of human resource management work, but also conducive to the targeted arrangement of training work.

2. BASIC PRINCIPLE OF DECISION TREE ALGORITHM

A decision tree is a predictive model that represents a mapping relationship between an object's properties and its values. Each node in the tree represents an object, each bifurcation path represents a possible attribute value, and each leaf node represents the value of the object represented by the path taken from the root node to that leaf node.

The machine learning technique of generating decision trees from data is called decision tree learning, or decision tree in popular parlance.

A decision tree contains three types of nodes:

Decision nodes: Usually represented by a rectangular box.

Opportunity nodes: Usually represented by a circle.

Terminal nodes: Usually represented by triangles.

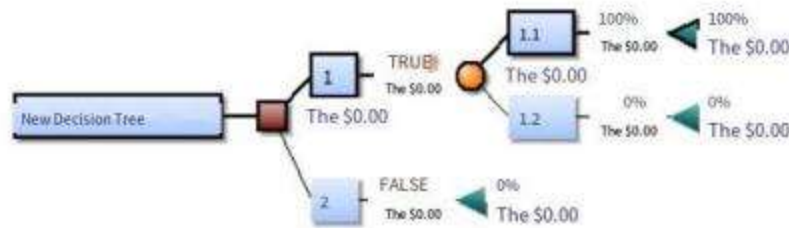


Figure 1. Decision tree view

The decision tree is a tree-like structure, in which each leaf node corresponds to a classification, and the non-leaf node corresponds to a division on a certain attribute, and the sample is divided into several subsets according to different values on the attribute. For non-pure leaf nodes, the labels of most classes give the class to which the samples arriving at this node belong. The core problem of constructing the decision tree is how to choose the appropriate properties to split the sample at each step. For a classification problem, it is a top-down, divide-and-conquer process to learn and construct a decision tree from training samples with known class markers.

3. IMPROVED XGBOOST DECISION TREE ALGORITHM

Most employment prediction systems are built decision tree models[3], this article builds the XGboost decision tree with data from a job Posting website.

XGBoost is an objective function calculation method based on gradient lifting, which improves the accuracy of the model, converts the solution method of the objective function into a quadratic function to solve the minimum value problem, and adds the complexity of converting the regular term into the tree model into the objective function, improving the computing power and generalization ability of the model. The modeling process is as follows:

- (1) Index selection, using literature research method, case analysis method to predict the salary index screening, and extract the sample of the corresponding index as the basic data of the prediction model construction.
- (2) In the data pre-processing stage, the selected salary prediction indicators were processed by multicollinearity and standardization, and outliers were eliminated to reduce the influence of outliers on the prediction accuracy of the model.
- (3) After data preprocessing, the data is divided into test set and training set to prepare for the normal operation of the model.
- (4) Parameters of the XGBoost algorithm model are optimized by using cross-validation and grid search methods, and the optimal combination of parameters of the prediction model is trained in the training set.
- (5) Test the model parameters selected in step (4) by using test samples, evaluate and analyze the results of the test set by using multiple evaluation indicators, and then randomly select sample data to verify the prediction of the XGBoost algorithm model and test the accuracy and stability of the model prediction. The modeling process is shown in Figure 2.

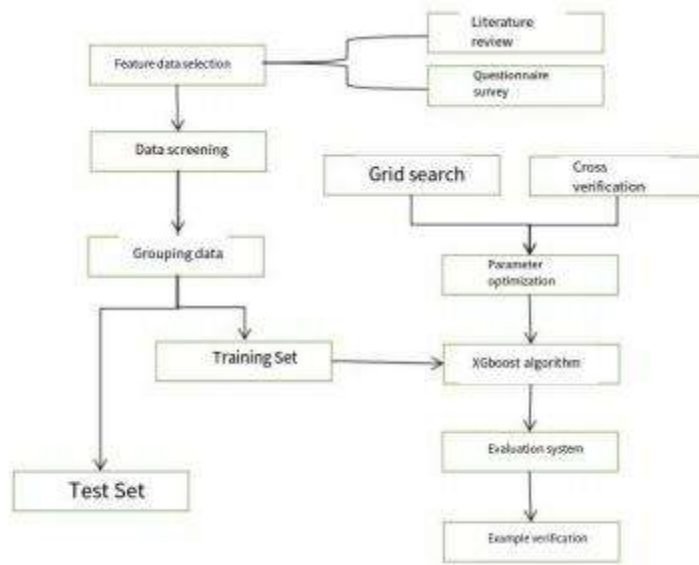


Figure 2. Modeling flowchart of XGboost algorithm

3.1. Job Application Requirements Analysis Function Module

According to the actual data, the work experience and education required by the post are divided and the results are calculated. Calculate the number and proportion of the job required to correspond to the work experience and the number and proportion of different academic qualifications, for fresh graduates and social personnel to understand the information of the target position is of great help. This paper uses XGboost decision tree algorithm, through the pie chart, line chart and other forms of the enterprise released salary, enterprise situation, company size distribution, education and work experience distribution were detailed visual display [4], so that the effect of data analysis can be improved. The analysis of job application is shown in Figure 3.

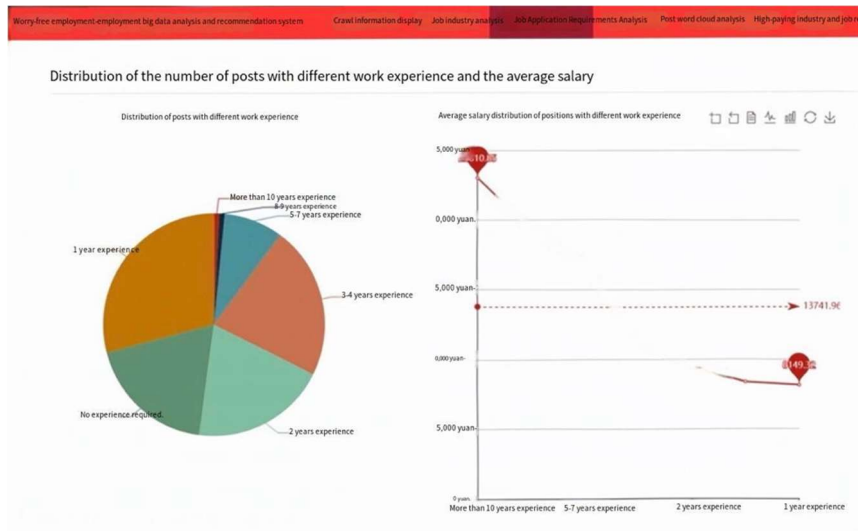


Figure 3. Analysis chart of job application requirements

3.2. Job Application Requirements Analysis Function

This paper uses web crawler technology, through certain rules to write code, automatic extraction of web information program, the entire web page can be completely crawled down, you can get detailed data and graphics, and has a high degree of customization. The core code is shown in Figure 4.

4. CONCLUSION

In order to solve the problems of college students' job hunting and employment difficulties, this paper establishes a decision tree model and proposes an improved Xgboost decision tree algorithm. The use of big data technology to analyze, process and visualize job data can not only make students more intuitive to understand the current employment market and demand, but also let the professional person in charge and the school educational administration department get the latest job positioning, job ability demand and other information. It can help users clearly and intuitively see the relationship between supply and demand, quickly find the recruitment information that meets their needs, and provide a convenient, accurate and fast intelligent post analysis platform for the majority of job seekers.

ACKNOWLEDGEMENTS

This work was supported by the fundamental research funds for Ministry of University-Industry Collaborative Education Program (220905181091456), Innovation and Entrepreneurship Training Program of Wuhan Business University (202211654165).

REFERENCES

- [1] Yuting Ye. Where do famous graduates find jobs? *Modern Youth*, 2022, (03): 38 -40.
- [2] Lei Hao. The application of job analysis in enterprise Human resource management [J]. *Human Resources Development*, 2023, (12): 71-73.
- [3] Mingguang zhang. Ru Ning. Research on Supply and demand matching of college graduates' employment under the background of Industrial transformation and upgrading [J]. *Exploration of Higher Education*, 2020(9): 114-122.
- [4] Liang Chen. Design and Implementation of Intelligent Job Analysis System based on Collaborative Filtering Algorithm [J]. *Software Engineering*, 2023, 26 (10): 58-62.
- [5] Zhiyang Song, Zhou Chen, Zhang Juan, Chen Yingying. Design and implementation of Job Analysis System for Recruitment Website [J]. *Computer & Telecommunications*, 2020, (09): 6-10.