Job Industry Analysis based on Decision Tree Algorithm

Yunyang Liao, Yunxiao He, Juanjuan Mu, Jiawen Chen, Juan Li

Wuhan Business University, Wuhan, China

Abstract: Influenced by various factors such as market economic policies, the employment problem of college graduates has become more and more significant. Therefore, the design of employment job industry analysis based on decision tree algorithm is significant for job seekers. Aiming at the influencing factors such as region, specialty, salary and so on, combined with the prediction model data preparation, data pre-processing, correlation analysis of data, visualization technology, to realize the job industry analysis based on the decision tree algorithm, to provide certain reference basis for the job seekers in the job-seeking project.

Keywords: Data Visualization, Decision Tree Algorithm, Post Data Analysis, Data Preprocessing, Word Cloud Analysis.

1. Introduction

In recent years, with the development of the economy and technological progress, more and more industries are changing, and with the increase in the number of college graduates, the employment of college students has become very difficult, so job industry analysis has become more and more necessary. Job industry analysis has an important role for enterprises and individuals. For enterprises, job industry analysis can help them understand the market trend and competition pattern, so as to make more effective strategies and plans. In addition, by analyzing the flow of talent within the industry, companies can better recruit and train talent to improve their competitiveness and performance. For individuals, job industry analysis can help them better understand the job market and career trends so that they can make more informed career choices. By understanding the working environment and treatment within the industry, individuals can better plan their career development and improve their career competitiveness.

To summarize, job industry analysis has become increasingly important in the current society. It can help enterprises and individuals better understand the market trends and competitive landscape, so as to make more informed decisions and planning. Therefore, this paper focuses on the research and application of job industry analysis, will crawl from the major job boards to pre-process and categorize the information, the processed recruitment data imported into the database, extract the number of industries and the number of specific people needed for each recruitment position, the use of decision tree algorithms modeling, through the statistics to get the number of people need to be recruited in different industries and the proportion of them. Finally, the pie chart form to the user to visualize. It provides an important reference for enterprise development and personal employment, so as to better cope with future challenges and opportunities.

2. Introduction to Decision Tree Algorithms and Models

A decision tree is a predictive model that represents a mapping relationship between object attributes and object values. Each node in the tree represents a particular object, while each forked path represents a possible attribute value, and each leaf node corresponds to 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, commonly known as a decision tree. A decision tree contains three types of nodes: decision nodes: usually represented by rectangular boxes; opportunity nodes: usually represented by circles; and end nodes: usually represented by triangles.

A decision tree is a tree-like structure in which each leaf node corresponds to a classification, and a non-leaf node corresponds to a division on an attribute that divides the samples into a number of subsets based on their different values on that attribute. For non-pure leaf nodes, the majority class labeling gives the class to which the samples arriving at this node belong. The central problem in constructing a decision tree is how to choose the appropriate attribute to do
the splitting of the samples at each step. For a classification problem, learning and constructing a decision tree from training samples with known class labeling is a top-down, divide-and-conquer process.

The advantages of the decision tree algorithm are as follows:
1. High classification accuracy;
2. Generate a simple model;
3. Good robustness to noisy data.

It is thus one of the most widely used inductive inference algorithms in the data mining in which it has received extensive attention from researchers.

Decision Tree as a supervised learning model in machine learning, which is concise and efficient, is a training structure algorithm with excellent performance and more mature development. Decision Tree (Decision Tree), also known as decision tree, is an important classification and regression method in data mining technology, which is a predictive analysis model expressed in the form of a tree structure (including binary trees and multinomial trees).

3. Job Industry Analysis

The function is to analyze the recruitment information of the job site to analyze the statistics, the data will be processed and imported, through the statistical function, the recruiters will be divided into industries, the statistical results, calculate the number of people recruited and the proportion of jobs in various industries, shown in the form of pie charts, calculate the average salary of different industries, and shown in the form of bar charts, which is very convenient for job-seekers looking for a job. This function will analyze the job industry on the employment data, you can get detailed data as well as graphs, analyze the job industry to get better data and charts, job industry analysis as shown in Figure 2.

![Figure 2. Job Sector Analysis Chart](image)

**Figure 2. Job Sector Analysis Chart**

Hot jobs word cloud analysis test: In this function display, the use of word cloud can be visualized to see the hot jobs in different regions, as shown in Figure 3.

![Figure 3. Word Cloud of Hot Jobs in East China](image)
By Figure 3 of the East China hot jobs word cloud map we can clearly see that the East China engineers hot jobs ranked first, and the number of jobs is 21045; sales jobs hot jobs ranked second, and the number of jobs is 17511; computer software jobs ranked third, and the number of jobs is 10640; manager jobs hot jobs ranked fourth, and the number of jobs is 9590; Heavy industry jobs Hot recruiting ranked fifth, and the number of jobs is 9127; students who want to go to East China to develop, learning these majors have great advantages.

By Figure 4 of the hot job word cloud in North China we can clearly see that the hot recruitment of sales positions in North China ranked first, and the number of jobs for 1023; the hot recruitment of engineers ranked second, and the number of jobs for 950; the operation of jobs ranked third, and the number of jobs for 827; commissioner positions ranked fourth, and the number of jobs for 744; logistics positions ranked fifth, and the number of jobs for 707; students who want to go to the development of these majors will have a great advantage. The number of positions is 707; students who want to develop in North China will have a great advantage in studying these majors.

4. Conclusion

The use of decision tree-based job industry analysis to analyze and guide the employment situation of college graduates, and targeted development of employment guidance programs, which is conducive to the job search of graduates, and conducive to improving the employment rate of college graduates. Colleges and universities can promote the employment policy, employment situation, job industry analysis, etc., to promote college graduates to understand the employment situation, in order to find the preferred job to improve their employment resilience and professional ability, and reasonably set the employment goals, to ensure the quality of employment. In general, decision tree-based job industry analysis has good application value.

Acknowledgments

The authors gratefully acknowledge the financial support from Innovation and Entrepreneurship Training Program of Wuhan Business University (202311654190), Ministry of Education Industry-University Cooperative Education Project (220905181091456).

5. References


