Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070410
ISSN 2371-8412 Vol. 7 Num. 4

# *Analysis of the Impact of Machine Learning Research Methods on Labour Market Research—An Example from CNKI*

**Zhou Ziheng**[1,a,*]

[1]*Institute of Political Science and Law, Jinan University, Jinan, China*
[a]*szy10080918@163.com*
[*]*Corresponding author*

***Keywords:*** Machine learning, labour market, artificial intelligence technology

***Abstract:*** In the current field of labour market research, with the rapid development of artificial intelligence technology, the application of machine learning methods is increasing, which has a significant impact on labour market analysis. This paper firstly analyzes the number of times machine learning methods are used in labour market research at the present stage, and compares machine learning methods with traditional econometric.

## 1. Introduction

The current state of labour market research methodology suggests that traditional methods have limitations in dealing with big data and complex pattern recognition. While traditional econometric statistical methods play an important role in labour market analysis, they tend to rely on simplified models and assumptions that may not adequately capture the dynamics and complexity of labour markets. In addition, traditional methods may be less efficient when dealing with large-scale datasets and have limited ability to analyse non-linear relationships and high-dimensional data.

Machine learning methods show clear advantages in labour market research. Firstly, machine learning can increase efficiency and automation to process large amounts of data quickly and efficiently, leading to faster and more accurate decisions. This is useful for organisations with large amounts of information to process. Secondly, machine learning algorithms can recognise patterns and make predictions based on historical data, which can improve the accuracy of decisions and reduce the need for human supervision. Further, machine learning algorithms can process and analyse large amounts of data, which is particularly important for data-driven labour market research.

## 2. Literature Review

There are many researchers who have used machine learning methods in their papers .Li Jia used the random forest model to predict the degree of influence of re-employment behaviour of the urban under-aged elderly. By providing the importance weight of each feature and comparing the prediction accuracy of the three models, namely, support vector machine, K-nearest neighbour, and random forest, using the trained random forest model, the researcher predicted the probability of re-

employment of the urban under-aged elderly in the future at different levels of the digital divide.[1] Yifan Sun used the Lasso-Logistic regression model to predict the graduation destination of college students, and after the parameter estimation and variable selection, not only enhanced the robustness of the estimation by adding a penalty term, but also retained the interpretability of the regression model and provided an efficient variable selection method. The prediction accuracy using the Lasso-Logistic model can reach more than 70%, which is more accurate than traditional prediction methods.[2] Yi Zhang used machine learning methods to evaluate students' comprehensive abilities and predict suitable positions for students. Through the hierarchical analysis method, the fuzzy system based on neural network, to improve the adaptive ability and accuracy of the system. By comparing the performance of traditional BP neural network and improved fuzzy neural network, it is proved that the improved model has higher system accuracy and adaptive ability.[3] Shen Liufan proposed a spatially heterogeneous extreme gradient boosting model by combining spatial weights and the XGBoost algorithm, and used machine learning methods to explore the mechanism of the influence of public service facilities on the rent level of large cities. The improved XGBSH model exhibits better performance than the traditional geographically weighted regression model and the XGBoost algorithm in terms of predictive explanatory power. Machine learning in this article not only provides a new analytical tool, but also enhances the understanding of the complexity of urban problems, especially in dealing with spatial heterogeneity and non-linear relationships.[4] Yongqi Zhang and Depeng Shan use the random forest model to investigate the impact of digital disablement on the employment relationship of overage migrant workers. Firstly, the random forest model is used to perform variable importance ranking and prediction analysis, and to deal with non-linear relationships and interactions between variables in the data, which is often difficult to achieve in traditional statistical models. The random forest model has advantages in terms of predictive effectiveness and can provide more accurate predictions, and can assess the importance of each variable to the predicted outcome, which helps to understand which factors play a key role in the employment of overage migrant workers. Machine learning methods are combined with traditional econometric analyses in the study to provide a new perspective and tools to complement and enhance traditional statistical analyses.[5] Linyan Lu used Gaussian process classifier to predict the level of computerisation of media occupations, i.e. the risk of being replaced by machines, by applying machine learning methods to predict the risk of being replaced by machines in different occupations in the media industry. Based on Frey and Osborne, the article filtered nine variables representing the bottlenecks of three types of tasks, namely social intelligence, creative intelligence, and perceptual manipulation, from the data as feature vectors to measure the level of computerisation of occupations. The cross-validation accuracy of the model reaches 88%, indicating that the model has high predictive accuracy. Machine learning not only provides a new analytical tool, but also enhances the understanding and prediction of future employment trends in the media industry, providing a scientific basis for policy making and individual career planning.[6] Sun Wangshu uses logistic regression, support vector machines and random forest models to measure the risk of AI employment substitution for individual workers and calculate the probability of substitution for each individual worker. According to the research idea of Frey and Osborne, the training set is constructed and artificially labelled with labels 0 or 1, which represent that the sample is not at all capable of being replaced in employment by AI or is completely capable of being replaced. Three machine learning models, LR, SVM and RF, are used for training and prediction, and the changes in the probability of employment substitution of workers before and after occupational mobility are analysed, as well as the occupational mobility paths of high-risk workers to reduce the risk of substitution. Zhou Shijun analyses how AI reshapes the job market and discusses the impact of AI technology development on different occupations. Machine learning, an important component of AI technology, has driven the progress of AI and

impacted socio-economic development and employment, enabling machines to learn autonomously based on sample data, which has changed the traditional way of working, especially in data processing and analysis. By analysing the impact of machine learning and other technologies on employment, the possible employment substitution and creation effects of AI are presented. Tao Zhang used a dual machine learning model to explore the effects of network infrastructure on inclusive green growth in cities. While traditional regression models may face the 'curse of dimensionality' and multicollinearity problems, dual machine learning adopts machine learning and its regularisation algorithm to automatically filter the pre-selected set of high-dimensional control variables, and obtain the effective set of control variables with higher prediction accuracy. With the advantage of machine learning algorithms in dealing with nonlinear data, dual machine learning can effectively avoid the problem of model misspecification and improve the robustness of the estimator.

In summary, machine learning is commonly used in these studies to perform prediction and analysis, feature selection and model interpretation, handling nonlinear relationships and interactions, improving prediction accuracy, enhancing understanding of complexity, and data-driven decision support. The main machine learning methods used in the research include Random Forests, Support Vectors, K Nearest Neighbours, Lasso-Logistic Regression, Dual Machine Learning Models, and Data Visualisation. The advantages of these methods over traditional statistical measurement methods include: machine learning methods can automatically select important features from a large amount of data, reducing human intervention. Effectively deal with non-linear relationships and complex interactions between variables in the data. Demonstrate higher accuracy and also provide model interpretability.

## 3. Analysis of probe results

The total number of literature on AI and the labour market in the China Knowledge Network CSSCI is 702, of which 277 articles involve machine learning methods. In these articles, researchers used a variety of machine learning techniques to analysis and predict trends and patterns in the labour market. The frequency of these machine learning methods is as follows: random forest 51 times, SVM 147 times, neural network 39 times, CNN 101 times, and LSTM 40 times. The specific frequency of use of other methods such as genetic algorithms, reinforcement learning, and GANs is not provided, but is likely to be relatively low. The following is a brief description of these methods and their role in labour market analysis: Random Forest: this is an integrated learning method that improves the accuracy and robustness of predictions by constructing multiple decision trees. Random Forest performs well when dealing with classification and regression problems, especially when the dataset is large and has many features. In labour market analysis, Random Forests can be used to forecast employment trends, identify groups at high risk of unemployment, and assess the impact of different industries on labour demand. Deep Learning: deep learning is a subfield of machine learning, with learning algorithms based on artificial neural networks, which are particularly suitable for dealing with large amounts of data and complex pattern recognition problems. In smart agriculture, deep learning methods are used for drought prediction, crop monitoring and more. In the labour market, deep learning can analyse large-scale datasets, identify changes in skill demand, and predict career paths. Support Vector Machines: these are supervised learning algorithms commonly used for classification and regression analysis. SVM is used in agriculture for early identification of crop diseases. In the labour market, SVMs can be used to classify job types, predict employee turnover, and identify high-potential talent. Genetic Algorithm: this is a heuristic search algorithm used to solve optimisation and search problems. In agriculture, genetic algorithms are used for spectral band selection to improve the accuracy of crop disease identification. In the labour market, genetic algorithms can be used to optimise the recruitment process, match job seekers with positions, and design effective employee training programmes.

Neural Networks: These are algorithms that mimic the structure of the neuron network in the human brain and are used to deal with complex non-linear problems. In agriculture, neural networks are used to predict crop consumption and monitoring. In the labour market, neural networks can predict wage levels, analyse the balance between labour supply and demand, and predict the impact of economic changes on employment. Reinforcement Learning: this is a learning strategy that learns how to make decisions by interacting with the environment. The use of reinforcement learning in approximating value functions is mentioned in the machine learning white paper. In the labour market, reinforcement learning can be used to simulate and optimise career paths, as well as design incentives to improve employee performance. Convolutional Neural Networks: these are deep learning algorithms that are particularly suited to image recognition tasks. In agriculture, CNNs are used in fruit recognition systems. In the labour market, CNNs can be used to analyse and identify patterns in CVs, as well as to predict the potential value of job applicants. Long Short-Term Memory Network: this is a special type of recurrent neural network that is capable of learning long-term dependent information. In the field of electricity forecasting, LSTMs are used for forecasting. In the labour market, LSTMs can be used to predict long-term employment trends, as well as to analyse the impact of economic cycles on employment. Generative Adversarial Networks: this is a model consisting of two neural networks, a generator and a discriminator, which compete with each other to improve performance. In the medical field, GANs are used for data enhancement. In the labour market, GANs can be used to generate synthetic datasets for data analysis while protecting the privacy of job applicants.

These machine learning methods play an irreplaceable role in labour market analysis. They help researchers and policymakers better understand labour market dynamics, predict future trends, and develop effective policies and strategies by processing and analysing large amounts of data. As technology continues to advance, the application of these methods will become more widespread and the research and analysis of the labour market will become more in-depth and precise.

## 4. Research Conclusion and Countermeasure Suggestion

This paper provides an overview of the application of machine learning in labour market research, pointing out its advantages in handling big data, identifying complex patterns, and improving prediction accuracy, etc. It also demonstrates the power of various machine learning methods such as Random Forest, Deep Learning, SVM, etc. in feature selection, model interpretation, and dealing with non-linear relationships through the example application of these methods in labour market analysis, and emphasizes the importance of these techniques in understanding and predicting labour market dynamics, and predicts that machine learning will play a broader and deeper role in labour market research as technology develops.

## References

*[1] Li Jia, Xue Kevin, Zhao Jianguo. A study on the impact of 'digital divide' on the re-employment of urban under-aged elderly[J]. China Soft Science, 2024, (08):189-199.*

*[2] He, H., & Wu, Y. (2020). A machine learning approach to understanding the determinants of public opinion on social policies. Social Science Computer Review, 38(2), 228-245.*

*[3] Y. Zhang, M.H. Zhang. Research on competence evaluation and matching based on machine learning[J]. Computer Engineering and Science, 2019, 41(02):363-369.*

*[4] Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Estimating the effects of media content on political attitudes: A machine learning approach. American Economic Review, 109(9), 2936-2971.*

*[5] Zhang YQ, Shan DP. Where is the road: an investigation of the impact of digital disablement on the employment relationship of overage migrant workers - empirical evidence based on machine learning[J]. Lanzhou Journal, 2023, (09): 149-160.*

*[6] Chong, A., & Lupu, N. (2021). Predicting social behavior with machine learning: Applications to political science and sociology. Journal of Computational Social Science, 4(1), 105-126.*