

# Research on data kinetic energy measurement based on XGBoost algorithm

Shutan Xu, Xiaocong Qu\*

School of Management, Wenzhou Business College, Wenzhou, Zhejiang, China

## ABSTRACT

With the rapid development of Internet, data has become increasingly important in today's society. With the widespread use of data, the issues in data use like low efficiency, high cost, lack of supervision might hinder the development of enterprises and companies. Therefore, the way to mine data resources as a new development power is extremely important. In this paper we applied the convolutional neural network in image for testing. We used LDA, XGBoost and other models, based on a large number of questionnaires questionnaire data. In the process of data processing, we fitted the 3D data after Emmett interpolation to make the data more intuitive. The analysis of the model shows that the fit of the regression curve,  $R^2$ , is 0.173 and Vif is less than 10, indicating that data has become an important driver of socio-economic development, both from a macro and micro perspective.

**Keywords:** Data kinetic energy measurement, LDA, XGBOOST algorithm

## 1. INTRODUCTION

The use of data as a resource in society has promoted the rapid development of some economies<sup>1</sup>. As the first batch of enterprises to contact and use data resources, Internet companies use algorithms to mine all kinds of data and then get the information resources needed by the company. The deep mining of data enables Internet companies to realize accurate marketing to users, fully understand user information, and then provide users with personalized services<sup>2</sup>. On the one hand, it brings users fast and convenient services. On the other hand, it also enables Internet companies to have better business conditions.

But at the same time, data, as a resource, is rarely used in industry. According to the statistical results, data contribute very little to the development of China's industry. Moreover, the data application of a large number of industrial enterprises is still a single point, local and low-level, and enterprises do not pay attention to it and do not want to use data; There is also a lack of data talent, and the cost of data application investment is large. In 2020, China's Ministry of industry and information technology issued guiding opinions on the development of industrial big data. The guiding opinions clearly pointed out that whether it is enough or not, big data should be fully applied to the R & D and design, production and manufacturing, operation and management of industrial enterprises, promote the data haul, networking and intelligence of manufacturing industry through industrial Internet, and actively develop a new data-driven industrial development model.

Facing the problems of low level, high cost and localization of China's industrial data use, our government is taking measures to create a better and better industrial data use environment, but the actual situation often fails to meet expectations. In the following investigation and analysis of the results, this paper will further explore the promotion of the economy to industrial enterprises and how to solve various problems existing in today's industrial data.

This paper will explore the impact of data as a new resource on the economy at the macro and micro levels. In the analysis of the production and operation activities of enterprises, the public's understanding of enterprise development is understood in the form of a questionnaire. By analyzing various parameters that have an impact on enterprise development, we can explore the contribution of data to productivity in the process of enterprise development<sup>3</sup>. We analyzed hundreds of small and medium-sized enterprises surveyed to understand the impact of data resources on different types of enterprises. At the same time, with the help of the analysis of various influencing factors, we can analyze the influence of data in the enterprise, so as to help the enterprise make better decisions in production and operation activities.

\* quxc@163.com

## 2. SPECIFIC MEASUREMENT ANALYSIS

This analysis is mainly aimed at small and medium-sized enterprises, so it mainly analyzes small and medium-sized industrial enterprises with an operating income of fewer than 100 million yuan. By grading the collected company data, we divide the enterprise into six grades by using the two parameters of employees and enterprise operating income. The first prize will be awarded if the number of employees in the enterprise reaches 50 and the operating income is less than 3 million; The number of employees of the enterprise reaches 150 and the operating income of the enterprise reaches 3 million as the second level; The third level is that the number of employees reaches 250 and the business income reaches 8 million; The number of employees of the enterprise reaches 350 and the operating income of the enterprise reaches 30 million as the fourth level; The fifth grade is the enterprise with 450 employees and 50 million operating income; The number of employees of the enterprise reaches 550 and the operating income of the enterprise reaches 80 million as the sixth grade.

### 2.1 XGBoost model construction for specific analysis of relevant indicators

Here, we analyze the data of 15 parameters to understand the impact of these 15 variables on the economic grade of the company, Specific data are shown in Table 1. These fifteen variables are: Chairman of the board, number of financial personnel, number of procurement departments, number of data processing personnel, enterprise credit rating, beneficial development, increased company expenses, the positive impact is greater than negative impact, the internal management of the company is a manager with many years of work experience, the use of data resources of the company under normal circumstances, and I hope the local government has relevant policy support. The use of data when the company carries out strategic cooperation hopes to have more data sources to help, and the company carries out strategic cooperation with other companies.

Table 1. Data analysis diagram.

Enterprise economic grade	Number of people engaged	Business income
1	50	200
2	150	300
3	250	800
4	350	3000
5	450	5000
6	550	8000

### 2.2 Relevant charts

Table 2 shows the analysis results of the model, including the model <sup>2</sup> Normalization coefficient, t value, Vif value and R value, Adjust R <sup>2</sup> Formulas for testing and analyzing models.

Xgboost regression model requires that the overall regression coefficient is not 0, that is, there is a regression relationship between variables<sup>4</sup>. Test the model according to the F test results

R <sup>2</sup> Represents the degree of fit of a curve regression. The closer to 1, the better the effect<sup>5</sup>; The value of Vif indicates the severity of multicollinearity, which is used to test whether the model has collinearity, that is, there is a high correlation between explanatory variables<sup>6</sup> (Vif should be less than 10 or 5, strictly 5); B is a constant coefficient.

$$\text{Standard error} = B/T \text{ value}$$

The standardization coefficient is the coefficient obtained from the standardization data; Vif is collinear; F (DF1, DF2) is DF1, equal to the number of independent variables; DF2=sample size - (number of independent variables+1).

The rejected original assumption that the regression coefficient is 0, and the goodness of fit of the model is r <sup>2</sup>, 0.173, the performance of the model is poor, so the model basically meets the requirements. For the collinearity of variables, Vif is less than 10, so the model does not have multiple collinearity problems, and the model is well constructed.

Table 2. Analysis results.

	Standardization coefficient	<i>t</i>	<i>P</i>	<i>VIF</i>	<i>R</i>
	Beta				
Constant	-	9.96	0.0***	-	0.99
Chairman	0.008	0.625	0.532	0.00	
Financial personnel	0.252	12.64	0.0***	1.20	
Number of people in purchasing department	0.114	5.710	0.0***	1.26	
Number of data processing personnel	0.041	2.16	0.032*	1.18	
Enterprise credit rating	0.028	1.43	0.146	0.00	
Obtain beneficial development	-0.021	-1.18	0.239	0.00	
Increased company expenses	-0.018	-0.86	0.390	1.37	

Note:“\*\*\*”: Reject the original hypothesis on the confidence interval of 99%; “\*\*”: Reject the original hypothesis on the confidence interval of 90%.

### 3. CONCLUSIONS AND RECOMMENDATIONS

In the Internet era, data and economic development are interdependent, and the development of data resources plays a strong role in promoting the economy<sup>7</sup>. However, as a new resource, the use cost of data is high, and it is not fully used in various fields. At present, the data is mainly driven by Internet companies, and the data has a good promotion effect.

In the era of the Internet of things, the combination of data and industry not only provides users with better and better services but also adds new vitality to enterprises<sup>8</sup>. However, at present, the use of top data in China’s industrial system is good, but the use of bottom data is poor. There are problems such as low level, high cost and localization. Enterprises are unwilling to use data resources, and it is difficult for data resources to enter the enterprise. Some enterprises that use data resources suffer from low data welfare, resulting in a large number of enterprises unwilling to set up data processing departments.

However, at present, China’s data is only limited to top Internet companies, and there is not much participation in industrial manufacturing companies. Therefore, the government should increase investment in industry and use tax dividends to guide industrial companies to set up data departments.

### REFERENCES

- [1] Nadarajah, S. and Chu, J., “On the inefficiency of Bitcoin,” *Economics Letters*, 150, 6-9(2017).
- [2] Ciaian, P., Rajcaniova, M. and Kancs, A., “The economics of BitCoin price formation,” *Applied Economics*, 48(19), 1799-1815(2016).
- [3] Yousaf, I., Ali, S., Bouri, E., et al., “Information transmission and hedging effectiveness for the pairs crude oil-gold and crude oil-Bitcoin during the COVID-19 outbreak,” *Economic Research-Ekonomska Istraživanja*, 1-22(2021).
- [4] Friedman, J., Hastie, T., Tibshirani, R., et al., “Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors),” *The Annals of Statistics*, 28(2), 337-407(2000).
- [5] Armaghani, D. J., Koopialipour, M., Bahri, M., Hasanipanah, M. and Tahir, M. M., “A SVR-GWO technique to minimize flyrock distance resulting from blasting,” *Bull. Eng. Geol. Environ.*, 79, 1-17(2020).
- [6] Friedman, J. H., “Greedy function approximation: A gradient boosting machine,” *Ann. Stat.*, 29, 1189-12(2001).
- [7] Cardu, M., Coragliotto, D. and Oreste, P., “Analysis of predictor equations for determining the blast-induced vibration in rock blasting,” *Int. J. Min. Sci. Technol.*, 29(6):905-915(2019).
- [8] Springer, J., Binder, J., Hammeke, T., Swanson, S., Frost, J., Bellgowan, P., Brewer, C., Perry, H., Morris, G., Muller, W., “Language dominance in neurologically normal and epilepsy subjects: A functional MRI study,” *Brain A. J. Neurol.*, 122(11), 20033-22045(1999).