ISSN 2616-5767 Vol.4, Issue 4: 49-53, DOI: 10.25236/AJETS.2021.040409

Big data analysis of chicken market based on multiple evaluation scales

Shuqi Chen, Ran Zhao, Shuwen Ren

Qufu Normal University, Jining, Shandong, 273100, China

Abstract: *Focusing on the chicken market, this paper makes a big data analysis on the chicken sales parameters of ZhengDa Group. First of all, python software is used for data preprocessing, and the sales situation of ZhengDa chicken food is analyzed by descriptive statistics. In terms of sales, the commodity price and distribution of ZhengDa chicken products are analyzed in turn. Then use the comments of ZhengDa fried chicken chops to analyze the emotion, and establish a theme model to understand the overall feeling of users on food, so as to solve the problem of product shortage and improve customer satisfaction. Finally, time series analysis is used to predict the final data.*

Keywords: *chicken, big data, Sentiment analysis, time series prediction*

1. Introduction

In order to expand the production and supply of poultry meat, China's Ministry of rural agriculture will work out the "14th five year" development plan of animal husbandry, further promote and improve the policy support system of pig and broiler, optimize the regional layout, accelerate the industrial transformation and upgrading, continuously enhance the comprehensive production capacity and supply guarantee ability, and promote the high-quality development of pig and broiler industry [1].

In the past 60 years, China's chicken production has been increasing. The production and consumption of chicken in China is second only to pork, and China has become the second largest chicken producer in the world after the United States. However, in recent years, negative events such as "fast growing chicken", "h7n9 influenza", "Fuxi expired meat" have affected the consumption of chicken and its products, had a negative impact on the chicken market, caused a significant loss of economic benefits of the industry, and seriously affected its sustainable development. Therefore, this paper studies ZhengDa chicken, analyzes the sales volume, market performance and consumption of ZhengDa food chicken, so as to provide reliable market information for ZhengDa Company, and put forward reasonable suggestions for the opportunity, positioning and market promotion of chicken products in the retail channel of ZhengDa Company. How to develop the processing industry according to local conditions, change the consumption mode of live poultry, and make the variety, quantity and quality of poultry products better meet the market demand is an urgent problem for poultry industry and enterprises.

2. Data preprocessing and preliminary analysis

2.1 Data collection

First of all, the paper find the relevant information about ZhengDa chicken products on the Internet through python, including the name of the product, price, region, number of payers, number of comments, freight, unique ID of the product, store name, etc. Secondly, the evaluation of commodity 1 with higher sales volume is mined, including user name, evaluation time and evaluation content.

2.2 Descriptive statistics analysis of sales data

Descriptive analysis including the data central tendency analysis data, the frequency distribution of discrete degree analysis, data analysis, and descriptive analysis is to further data points.

(a) Situation analysis of the distribution of commodity prices

To understand the distribution of ZhengDa chicken products price, you can use the pretreatment data after drawing histogram [3]. Histogram drawing can use hist () function, you can also use the package

ISSN 2616-5767 Vol.4, Issue 4: 49-53, DOI: 10.25236/AJETS.2021.040409

ggplot2 for drawing the figure 1.

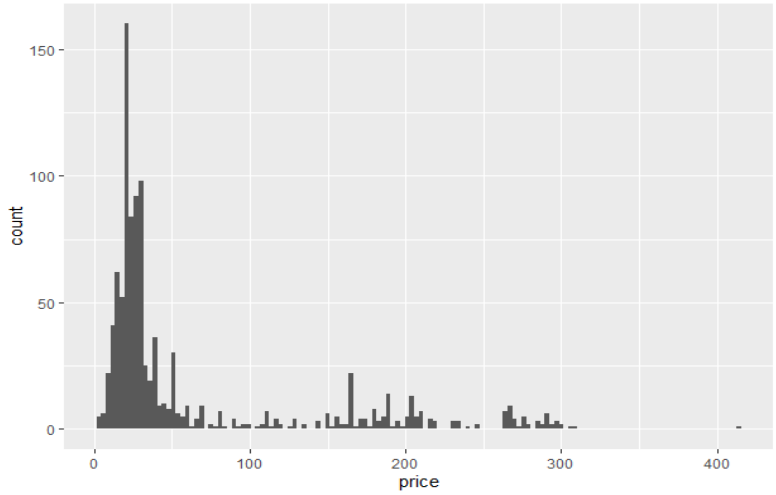

Figure 1: Commodities distribution histogram

The price distribution of the commodity exhibits a certain degree of right. The price is mainly distributed at 0-100 yuan, and the proportion of goods that exceed 100 yuan is very small. Chicken products are living consumables. Many customers will use them as part of their daily diet. Therefore, the price is controlled in a affordable reasonable interval, which is easier for customers to accept Master [4].

(b) Distribution of sales

The region column in the sales data is organized, and the distribution of the delivery address of the positive product can be obtained. Since the regionality is no longer a continuous random variable, use the R software to draw the column map of sales distribution. There is a certain gap between each of the columnar images, each of which represents a certain value level of the variable. The paper uses the histogram to draw the top ten regions of the product sales.

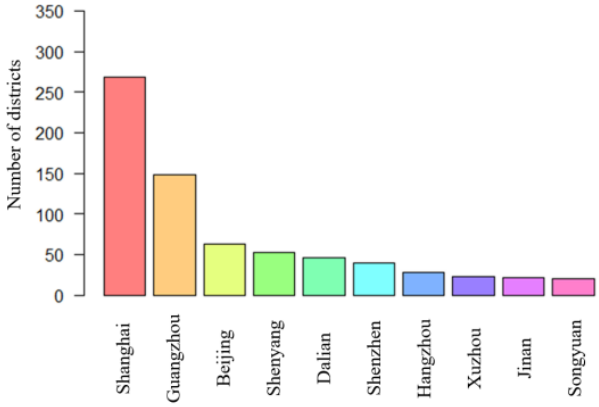

Figure 2: Regional ranking of goods sales

3. Sentiment analysis of comment model

3.1 Sentiment analysis

Emotional analysis is a common application of natural language treatment methods, especially in the classification method of extracting the emotional content of the text. It can be analyzed according to the hi, anger, sorrow, music, criticism, praise, etc. In people's comments, and draw each comment as the probability of positive emotions, the closer to 1, indicating the front emotions, the closer to 0, indicating negative emotions.

Snownlp is a Python written class library that can easily process Chinese text content. SnownLP is a model that uses Taobao's comment, using Bayesian machine learning methods. Using SNOWNLP, each comment is probability of positive emotions.

ISSN 2616-5767 Vol.4, Issue 4: 49-53, DOI: 10.25236/AJETS.2021.040409

Table 1: Each comment is the probability of positive emotions (part)

P	rate_content
0.94237	After many buybacks, children like to eat this bite, and they must buy it if they have any activities.
0.91792	It's delicious. It's fried in an air frying pan.
0.26778	It's delicious. It's been bought back for the second time.
0.94632	The packaging is exquisite, and the things are really authentic!
0.97631	Has been bought many times, the child's favorite, love, very delicious.
0.93782	Very fresh, the fastest delivery speed, thanks to the merchants have bought babies love to eat. It will be bought back in the future.

According to the probability of calculated positive emotions, the average value is divided into 5 parts, and the star is labeled: (0,0.2]-one star, (0.2,0.4]-two stars, (0.4,0.6]-three stars, (0.6,0.8]-four stars, (0.8,1.0]-five stars.

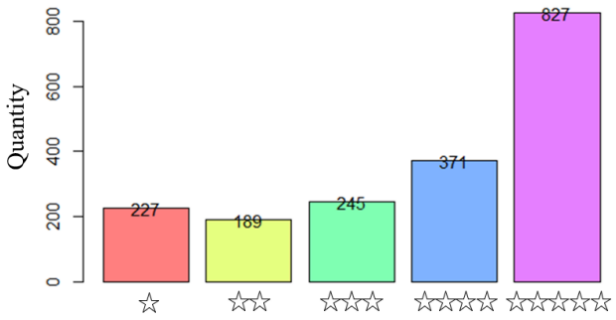

Figure 3: Commodity star statistics

According to the content of the picture, there are 827 comments on 5 stars, 371 comments on 4 stars, 245 comments on 3 stars, 189 comments on 2 stars and 227 comments on 1 star. 5 stars and 4 stars

Since there is a certain error in the star level of probability judgment obtained by SnowNLP analysis, the following analysis is carried out.

3.2 LDA theme model

The structure of the LDA model is divided into three layers, from top to bottom for documentation, theme, and words. The main idea is to treat the document as a probability distribution of multiple topics, regard each subject, the probability distribution of multiple words, expressed as the following probability formula [2]:

$$p\left(w_n \mid M_m\right)=\sum_{k\in K} p\left(w_n \mid K_k\right) p\left(K_k \mid M_m\right) \tag{1}$$

It is expressed as the probability of the occurrence of the word item *Wn* in the document M_m, the probability of the occurrence of the word item *Wn* in the topic K_k where $k \in K$. N is all lexical items, M is all documents, and K is all topics.

Analyze the forward comments (praise) and negative comments (bad reviews), and get the theme of each comment, which in turn overcomes certain aspects and improve customer satisfaction.

The positive and negative potential subject characteristic words are obtained respectively. Here, take the positive as an example. Positive words to the potential topic of the comments are the table:

Table 2: Positive comment 3 topics feature words

1	buy-back	recommend	texture	purchase	Blasting	Deep frying	Price	express	Cost effective	high
2	packing	logistics	Blasting	express	Praise	Ice bag	deliver goods	purchase	flavor	recommend
3	purchase	convenient	Blasting	deliver goods	Cost effective	stick	logistics	buy-back	recommend	speed

In negative comments, consumers are mainly concerned about product quality, including whether the taste of the product is in line with consumers, in addition to this, consumers are more concerned about the logistics speed of the product, because the logistics speed is also a major impact of determining product quality factor. Merchants should handle these two questions correctly and improve customer satisfaction.

ISSN 2616-5767 Vol.4, Issue 4: 49-53, DOI: 10.25236/AJETS.2021.040409

4. Sales analysis based on time series model

The monthly comment data approximates a monthly sales data, so the time series model can be established. Self-returning model AR(p) and mobile average model MA(q) can be considered as a special condition of the time series self-regression moving average ARMA (p,q) model, and the self-regression moving average model has the nature of both. Determine the order of the model and the model according to the self-correlation coefficient and the partial correlation coefficient. The model can be performed on the model after the model is determined, and the prediction of the time series is performed.

Since the sales volume of the genuine food is better, in order to establish a time series model, we need to climb a lot of data, and these comments began in August 2020, which in October 2020, you can see sales in October. It is also the highest, about 1575 products are sold. Statistical monthly data, draw a time series diagram.

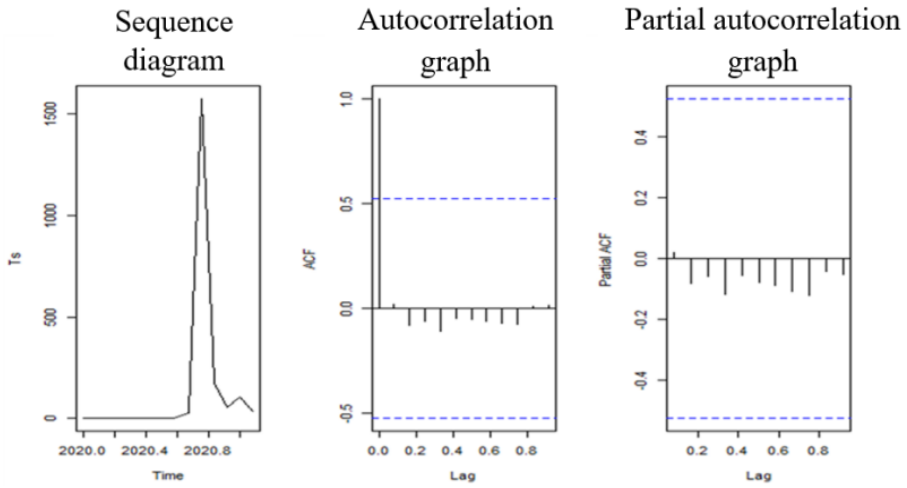

Figure 4: Time-sequence diagram, autocoric map, and partial correlation

According to the characteristics of the truncation, we can establish a first-order moving average model (1) mA, and the parameters can be estimated to get the model using the general likelihood method.

$$x_t = 140.0645 + \varepsilon_t - 0.0210\varepsilon_{t-1} \tag{2}$$

The significant test of the model is a white noise test for the residual sequence. *Lb(ljung-Box)* check statistic p is 0.9988, which is greater than the significant level 0.05, so it cannot refuse the original hypothesis, it is considered that the residual sequence is white noise.

The prediction of the model can be predicted using the Forecast function in the package Forecast, gives the predicted value, and gives 80% and 95% confidence intervals. As follows:

Table 3: Prediction and interval prediction

	Point prediction	Lo 80	Hi 80	Lo 95	Hi 95
2021.3	138	-375.7661	651.467	-647.6585	923.3601
2021.4	140	-373.6659	653.7948	-645.6184	925.7473

From Table 3, the number of comments in March this year is approximately 138, and it can be approximate that sales of sales in March is approximately 140 sales in September. The size of the confidence interval can reflect the predicted good, due to the characteristics of the time series, the model cannot be performed for long-term predictions. In addition, the time span of the comment is too small, starting from August 2020, in February this year, if there is a few years of sales data, the data is analyzed, the predictive error will be greatly reduced, and the future sales situation is also more reference value.

ISSN 2616-5767 Vol.4, Issue 4: 49-53, DOI: 10.25236/AJETS.2021.040409

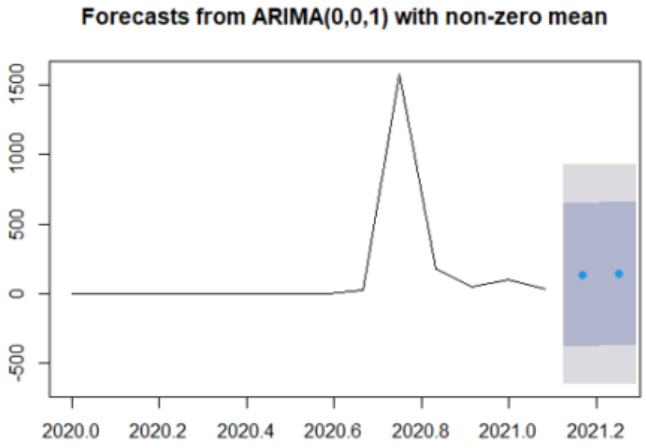

Figure 5: Point prediction and interval prediction

5. Conclusion

In this paper, the chicken market as the research object, the ZhengDa chicken sales parameters of big data analysis. A preliminary statistical analysis was made on the sales and price of ZhengDa chicken food. Then the paper use the comments of ZhengDa fried chicken chops for emotional analysis, establish a topic model, and finally use time series analysis to predict the final data. Provide new ideas for improving sales volume and satisfaction of commodity market.

References

[1] The replacement space of chicken increases, and the future market analysis [J]. Northern animal husbandry, 2020 (01): 6-7.
[2] Li Xiangdong, Huo Yayong, Zhang Jiao. Research on Extraction of bibliographic information from Book Web pages based on LDA topic model [J]. Information science, 2016, 34 (01): 34-37 + 43.
[3] Han Jing. Wheelchair network based on machine learning evaluation of sentiment analysis [D]. Lanzhou University of finance and economics, 2020.
[4] Lan Guicheng, Shan Yajun, Huang Jianqing. Dynamic interaction between online reviews and product sales: An Empirical Study Based on Taobao commodity panel data [J]. Journal of literature and data, 2020, 2 (03): 48-59.