Proceedings of the 3rd International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/34/20241167

An enhancement methodology for predicting transaction amounts for freelancers on freelance platforms: Based on sentiment analysis

Hongbin Zhang

School of Management, Zhejiang University, Zhejiang, Hangzhou, 310000, China

zhanghongbin20@zju.edu.cn

Abstract. Freelancing is a type of labor arrangement in which independent freelancers use their discretionary time to perform ad hoc tasks, usually of a single nature, with the aim of getting paid. This study proposes an augmented method for predicting the amount of freelance transactions using textual sentiment analysis. First, a unique feature labeled "Freelancer Sentiment" was selected to summarize the positive or negative sentiment orientation of freelancers. Subsequently, Naive Bayes algorithm is applied to process the text data from the freelancer's platform to finally develop a "Model for Computing Word Sentiment Values". The model helps to accurately calculate the sentiment values associated with emotional words. Finally, the Word Frequency-Inverse Document Frequency (TF-IDF) algorithm is used to construct the Text Sentiment Value Calculation Model, so as to accurately calculate the sentiment values of freelancers. The results of the comparison experiments of the five commonly used prediction models show that the mean squared error (MSE) of the model that includes the "freelancer sentiment" feature is significantly reduced by 6%-11% compared with the model that does not include the "freelancer sentiment" feature. This study contains theoretical explorations and practical implications. First, the proposed approach of extracting features from textual data to build predictive models provides a valuable reference for future enhancement of predictive modeling on freelance platforms, especially those that rely on unstructured data. Second, incorporating textual sentiment value features relevant to freelancers can significantly improve the accuracy of predicting transaction amounts. Third, the calculation of word and text sentiment values employs a series of algorithms that target specific features of the freelancer platform's text data. This approach is important for improving the accuracy of feature value calculation.

Keywords: freelance platforms, sentiment analysis, feature extraction, Naive Bayes algorithm.

1. Introduction

Freelancing represents a work arrangement where individuals, known as freelancers, dedicate their discretionary time to fulfill temporary tasks, often of a one-off nature, in exchange for compensation. The recent surge in internet platforms and the gig economy has played a pivotal role in advancing the development of freelance platforms [1]. As a result, a diverse array of freelance platforms has emerged, including prominent examples such as Freelancer, Elance, oDesk, and the Zhubajie website. These platforms establish a digital connection between demand and employer facets, thereby facilitating the execution of freelancing activities online.

DOI: 10.54254/2753-8818/34/20241167

Compared to traditional offline freelancing models, platform-based online freelancing transactions offer several advantages, including heightened openness, reduced transaction costs, minimal temporal and spatial constraints, and increased organizational process flexibility. Moreover, these platforms harness the power of big data for the recognition of demand and service features, customer-freelancer matchmaking, modularization, standardization, and the infusion of intelligence into demand and service processes. This integration enables demands to be addressed more conveniently and intelligently.

Within the freelance platform, addressing a diverse array of heterogeneous freelancers, the prognostication of their forthcoming transaction amounts grounded in their data characteristics is paramount for the platform. This is pivotal for the reasonable allocation of marketing resources, efficient organization of services, and attaining maximal business value [2].

In the present study focused on forecasting transaction amounts for freelancers on freelance platforms, the prevailing methodology centers around using structured data for predictive modeling. However, freelance platforms have witnessed a substantial influx of unstructured textual data, including user reviews and merchant self-introductions. This corpus of data encapsulates various dimensions, encompassing public sentiments regarding a public event [3], consumer perspectives on a specific product [4], user demands for a particular product [5], and user and employee satisfaction with a product [6]. Such information assumes paramount importance for the platform, facilitating the enhancement of operational strategies, formulation of informed business strategies, and refinement of specific products.

When consumers make choices among freelancers on freelance platforms, they often prefer individuals characterized by high levels of expertise and positive service attitudes [7]. Textual data within freelance platforms, such as freelancers' self-introductions and comments, encapsulate information that reflects freelancers' professionalization and patience levels. These factors have the potential to influence service quality and customer experience, subsequently impacting customers' willingness to make purchases and, consequently, the transaction amounts of freelancers. For instance, in this investigation, two self-introductions from freelancers on a reputable freelance platform were carefully selected (personal information has been anonymized), as depicted in Figure 1: The freelancer in (a) communicates a positive service attitude through an amicable language style, incorporating professional terminology and detailed service descriptions to instill confidence in customers regarding their business capabilities. In contrast, the freelancer in (b) employs random letters as placeholders, reflecting a perfunctory attitude that might prompt customers to abandon their intended purchases.

(a) (b)

Figure 1. Comparison of Freelancers' Self-introductions on a Freelance Platform (a) Self-introduction of a positive freelancer; (b) Self-introduction of a negative freelancer.

DOI: 10.54254/2753-8818/34/20241167

Current investigations into predicting transaction amounts for platform freelancers have generally neglected including unstructured data, specifically self-introductions. The process of extracting opinions, evaluations, attitudes, and sentiments expressed by individuals towards organizations, entities, individuals, issues, actions, topics, and their associated attributes from non-structured text is formally recognized as sentiment analysis [8]. Employing this methodology proves effective in facilitating the extraction of pertinent text features. In order to enhance predictive accuracy, considering the unique characteristics of textual data on freelance platforms, it is essential to successfully implement text feature extraction.

This study, grounded in the robust professional characteristics of textual data on freelance platforms [9], presents an enhanced methodology based on text sentiment analysis for transaction amounts prediction model for freelancers within the freelance platform environment. Initially, a distinct feature named " Freelancer Sentiment" is identified to capture freelancers' positive or negative sentiment values. Subsequently, the Naive Bayes algorithm is applied to process textual data sourced from freelance platforms, leading to the Word Sentiment Value Calculation Model's development for accurately computing sentiment words' sentiment values. Finally, the term frequency–inverse document frequency (TF-IDF) algorithm is employed to establish the Text Sentiment Value Calculation Model, ensuring precision in calculating the textual sentiment value, referred to as the Freelancer Sentiment Value. This study references the text features extracted through this method as " Freelancer Sentiment" in this study.

To substantiate the enhancement in predictive efficacy resulting from the extraction of text features in this study, we leveraged a dataset comprising both structured and unstructured data from 1306 freelancers on a freelance platform. We conducted comparative experiments employing five widely utilized prediction models, both with and without integrating the " Freelancer Sentiment" feature. The experimental outcomes demonstrated that, compared to prediction models lacking the " Freelancer Sentiment" feature, those incorporating this aspect displayed an average reduction in prediction error ranging from 6% to 11%. This corroborates the effectiveness of the proposed method in this investigation.

This study holds both theoretical and practical significance:

- The methodology introduced in this study, aimed at extracting predictive model features from textual data, offers valuable insights for advancing future predictive models on freelance platforms that rely on unstructured data.
- Incorporating the Freelancer Sentiment feature significantly enhances the precision of predicting transaction amounts for freelancers. This, in turn, assists platforms in devising judicious resource allocation and operational strategies.
- In the computation of word sentiment values and text sentiment values, this study utilizes a series of algorithms specifically designed to accommodate the unique characteristics of textual data within freelance platforms. This approach is paramount in elevating the precision of feature value calculations.
  - When calculating word sentiment values, this study, based on relevant texts from the freelance platform, employs the Naive Bayes algorithm to construct the Word Sentiment Value Calculation Model.
  - When computing the Freelancer Sentiment Value, this study integrates word frequency-adjusted weights and develops a Text Sentiment Value Calculation Model. This approach enables the precise calculation of the textual sentiment value, specifically referred to as the Freelancer Sentiment Value.

2. Literature review

Extracting opinions, evaluations, attitudes, and sentiments from the text of organizations, entities, individuals, issues, actions, themes, and their attributes is denoted as sentiment analysis [8]. Presently, predominant methodologies in sentiment analysis encompass lexicon-based sentiment analysis, machine learning-based sentiment analysis, and deep learning-based sentiment analysis. The overarching goal of

DOI: 10.54254/2753-8818/34/20241167

these methodologies is transforming unstructured textual data into structured formats, thereby facilitating the utilization of such data in predictive and classification tasks.

(1) Sentiment analysis based on the sentiment dictionary

The sentiment analysis based on a sentiment dictionary involves extracting and analyzing key sentiment words from the text to study the sentiment orientation of the text [10]. Currently, commonly used English sentiment dictionaries include WordNet [11], SentiWordNet [12], and LIWC [13], while popular Chinese sentiment dictionaries include NTUSD [14] and HowNet [15].

From the perspective of part-of-speech categorization, sentiment dictionaries can be classified into two types: single and multi-dimensional. Single-dimensional sentiment dictionaries involve the division of sentiments along a single dimension, with "positivity" being the most prevalent dimension [16 - 17]. However, a multi-dimensional sentiment dictionary becomes essential when the study goal is to comprehend the nuanced perspectives of a specific demographic on a particular event, product, or service. As exemplified Ahmed et al., they examined the sentiment orientation of comments on a video website, considering seven dimensions: disgust, sadness, like, anger, surprise, fear, and happiness [9].

(2) Sentiment analysis based on machine learning

Machine learning techniques are progressively gaining prominence in the domain of sentiment analysis. Prevalent methods in sentiment analysis employing machine learning encompass K-nearest neighbors (KNN) [18 - 19], support vector machines [18, 20], and Naive Bayes [21].

(3)Sentiment analysis based on deep learning

Deep learning is a category of machine learning models capable of automatically constructing features by establishing multi-layered neural networks for feature extraction from data. Deep learning methods find extensive application in the field of sentiment analysis. Commonly used deep learning methods for sentiment analysis include the BERT model [9], Convolutional Neural Network (CNN) [22 - 23], ELMo model [24], and Recurrent Neural Network (RNN) [24].

3. Feature extraction method for freelancers on freelance platforms based on sentiment analysis

This study introduces a method for extracting the Freelancer Sentiment feature based on text sentiment analysis. The derived feature, termed " Freelancer Sentiment," encompasses details regarding the freelancer's service demeanor, professional competence, and other factors influencing customer purchase intentions, thereby facilitating the prediction of freelancer sales. The fundamental procedure of this method includes (1) Feature selection, involving the definition of text features; (2) Application of the Naive Bayes algorithm to process text from the freelance platform, leading to the construction of the Word Sentiment Value Calculation Model for computing sentiment values of selected sentiment words; (3) Implementation of the TF-IDF algorithm to evaluate the significance of sentiment words, resulting in the establishment of the Text Sentiment Value Calculation Model and computation of the Freelancer Sentiment Value.

To mitigate ambiguity, this study terminology designates the sentiment value associated with sentiment words as "Word Sentiment Value," the sentiment value related to text as "Text Sentiment Value." It identifies the text sentiment value specific to freelancers as " Freelancer Sentiment Value."

3.1. Feather Selection

This study utilizes a single-dimensional (positive/negative) discrete sentiment dictionary, where the polarity of a word is either "positive (labeled as 1)" or "negative (labeled as 0)." Let the sentiment value of word i be denoted as sen_i, and the sentiment value of text j be denoted as $text\_sen_j$.

3.2. Construction of the Word Sentiment Value Calculation Model based on the Naive Bayes Algorithm.

The Naive Bayes algorithm is a frequently employed technique for computing word sentiment values. It determines the sentiment value of words by analyzing the frequency distribution of sentiment words in both positive and negative texts. Given its dependence on text sentiment values for calculating word sentiment values, it is particularly suitable for the specialized and domain-specific nature of sentiment words on freelance platforms [9]. Consequently, this study selects the Naive Bayes algorithm to formulate the Word Sentiment Value Calculation Model.

Initially, this study gathered textual data from widely utilized freelance platforms, comprising comments and self-introduction texts. Subsequently, this data underwent scoring based on the criteria of professionalism and positivity within the text. This process yielded 16,548 instances categorized as positive and 18,576 instances categorized as negative. The corpus underwent further analysis involving tokenization and word frequency computation, elucidating the frequency of each word's occurrence in both positive and negative datasets.

The frequency of word i in the positive corpus is f_pos_i, and in the negative corpus, it is f_neg_i.

Subsequently, a sentiment value calculation model is constructed using the Naive Bayes algorithm. The principle involves calculating the sentiment value of each word based on its frequency in both positive and negative corpora. For instance, the sentiment value of word i is:

$$sen_i = \frac{f\_pos_i}{f\_pos_i + f\_neg_i}, \quad (1)$$

Train the sentiment value calculation model with 4569 labeled positive samples and 4371 labeled negative samples from the corpus. Use accuracy, precision, and recall as metrics with the following calculation methods:

Accuracy = (Number of samples predicted as positive and are actually positive + Number of samples predicted as negative and are actually negative) / (Total number of samples)

Recall = Number of samples predicted as positive and are actually positive / Number of positive samples

Precision = Number of samples predicted as positive and are actually positive / Number of samples predicted as positive

The final classification results are presented in Table 1:

Table 1. Training results of the sentiment value calculation model on the corpus.

	Predicted positive words	**Predicted negative words**
Actual positive words	4079	490
Actual negative words	115	3523

Table 1 shows that the sentiment value calculation model achieved an accuracy of 85.0%, a recall of 89.3%, and a precision of 82.8%. Therefore, it is evident that the sentiment value calculation model, trained using the Naive Bayes model, demonstrates effective performance in classifying positive and negative sentiment words.

3.3. Construction of the Text Sentiment Value Calculation Model based on the TF-IDF Algorithm

In light of the potential significance of certain words with low occurrence frequencies in influencing the overall sentiment of the text, this study refrains from employing a straightforward weighted sum method for computing the freelancer's text sentiment value. Instead, it opts for the TF-IDF algorithm to develop the Text Sentiment Value Calculation Model, facilitating the computation of the Freelancer Sentiment Value. This methodology considers the frequency of word occurrence in the text and the general importance of each word [25].

The principle is as follows:

Let the importance of word i in text j be denoted as w_{ij}, and its relationship with word frequency is expressed as :

$$w_{ii} = \frac{f_{ij}}{\sum_{i \in I_j} f_{ij}} \times IDF_i, \quad (2)$$

DOI: 10.54254/2753-8818/34/20241167

Here, $\frac{f_{ij}}{\sum_{i \in I_j} f_{ij}}$ represents the frequency of word i in text j, IDF_i is the inverse document frequency (commonality) of word i, calculated as:

$$IDF_i = \log(N / (N_i + 1)), \tag{3}$$

The number of texts in the text set is represented by N, and the number of texts containing word i is represented by N_i.

Finally, the sentiment value of the text is calculated based on the sentiment values of the words, with the formula represented by:

$$text\_sen_j = \sum_{i \in I_j} \left(w_{ij} / \sum_{i \in I_j} w_{ij} \right) sen_i, \tag{4}$$

4. Validation

4.1. Introduction to the Dataset and Data Preprocessing

The experimental dataset utilized in this study consists of 1306 records of freelancers sourced from a widely employed freelance platform. Each entry encompasses not just the self-introduction of the freelancer but also incorporates 12 predictive indicators associated with the freelancer's prospective annual transaction volume. These indicators encompass attributes such as membership level, freelancer type, and creation time, as detailed in Table 2:

Table 2. Dataset Features and Meanings.

Feature Name	Meaning
Self-Introduction	Text provided by the freelancer during registration
Membership Level	Membership level of the freelancer on the platform
Freelancer Type	Whether the freelancer operates as an individual or a company
Creation Time	Number of days since the freelancer registered on the platform
Freelancer Scale	Categories include micro, small, medium, large, etc.
Times Favorited	Number of times the freelancer has been favorited by users
Number of Services on Sale	Number of services currently being offered by the freelancer
Showcase Cases	Number of transaction cases displayed by the freelancer on their profile
Total Historical Transaction Amount	Total transaction amount completed by the freelancer on the platform
Total Historical Transaction Volume	Overall transaction volume completed by the freelancer on the platform
Capability	ScoreNumerical assessment of the freelancer's capabilities by the platform
Employer Satisfaction	Satisfaction level of customers with the freelancer's transactions
Credit Rating	Credit score assigned by the platform based on the freelancer's transaction history

In terms of handling missing values, samples with a significant number of missing values (more than half of the features are missing) were discarded, resulting in 1016 remaining sample data. A "fill with 0" approach was applied for the remaining missing values. Regarding data processing, for categorical features like freelancer type, personal operation was designated as 0, and business operation was designated as 1. For the freelancer scale, the corresponding numerical values for micro, small, medium, and large were 0.25, 0.50, 0.75, and 1.00, respectively. For the credit rating, a percentage-based metric, considering that over 95% of the data was above 80, values below 80 were set to 0, and values greater than or equal to 80 were set to 1. For other numerical features, standardization and normalization were applied.

4.2. Comparison Methods and Evaluation Metrics

In forecasting the future annual transaction amount of freelancers, prevalent methods for predicting the price of a product, the transaction volume of a product, or the sales volume of a store encompass linear regression [26], regression trees [27], and ensemble learning techniques grounded in regression trees, such as random forests [28], GBRT [29], CatBoost [30], and so forth. This study opts for the five methods mentioned above as strategies for predicting the future annual transaction amount of freelancers.

The transaction amount is a type of continuous data, and Mean Squared Error (MSE) stands as a prevalent method for evaluating errors in continuous datasets. Consequently, this study uses Mean Squared Error (MSE) as the evaluation metric. The formula for calculating MSE is represented as:

$$MSE = \frac{\sum_{i=1}^{N}(f(x_i) - y_i)^2}{N-1} \tag{5}$$

In the context of the previous discussion, where N represents the number of samples, f denotes the prediction model, x_i represents the features of sample i, and y_i signifies the actual values of sample i, specifically the future annual transaction amounts.

4.3. Results

On the dataset, we employed five methods—Linear Regression (LR), Regression Tree (RT), Random Forest (RF), Gradient Boost Regression Tree (GBRT), and Categorical Boosting (CatBoost) —to predict the future annual transaction amount of freelancers. These predictions were conducted with and without including the Freelancer Sentiment feature. The dataset was partitioned into training and testing sets in an 8:2 ratio, and the results of ten-fold cross-validation are detailed in Table 3. The table illustrates that incorporating the Freelancer Sentiment feature led to an average reduction of 6%-11% in the errors associated with each prediction method. This enhancement is attributed to the Freelancer Sentiment feature effectively capturing information related to the freelancer's service attitude, level of professionalism, and other factors, thereby improving the accuracy of predicting the freelancer's future annual transaction amount.

Table 3. Comparison of Prediction Results with and without Freelancer Sentiment Feature.

Model	Without " Freelancer Sentiment" Feature MSE (in (hundred million CNY)2)	With " Freelancer Sentiment" Feature MSE (in (hundred million CNY)2)	MSE Change
LR	19.65	17.79	-9.48%
RT	25.48	23.45	-7.96%
RF	6.26	5.87	-6.23%
GBRT	1.98	1.78	-10.05%
CatBoost	1.94	1.77	-8.78%

5. Conclusion and future outlook

5.1. Conclusion

This study introduces an enhanced methodology based on sentiment analysis to predict the transaction amounts of freelancers on freelance platforms. To improve the precision of forecasting freelancer transaction amounts, the study follows these steps: Firstly, it selects the " Freelancer Sentiment" feature to capture freelancers' positive or negative sentiment values. Secondly, acknowledging the pronounced professionalism and domain specificity of data on freelance platforms [9], the study utilizes the Naive Bayes algorithm to handle text data on these platforms to construct the Word Sentiment Value Calculation Model to compute the sentiment values of sentiment words precisely. Lastly, considering the influence of word frequency and commonality on word importance, the study employs the TF-IDF algorithm to establish the Text Sentiment Value Calculation Model for freelancers, ensuring accurate computations of Freelancer Sentiment Value. Among the five common prediction methods, incorporating the Freelancer Sentiment feature results in a 6%-11% reduction in prediction errors compared to models without this feature.

This study encompasses both theoretical investigation and practical applicability. Firstly, the method delineated in this study for extracting features from textual data to construct predictive models serves as a benchmark for the prospective enhancement of predictive models within the domain of freelance platforms reliant on unstructured data. Secondly, incorporating the Freelancer Sentiment feature significantly augments the precision of predicting transaction amounts associated with freelancers. This advancement holds the potential to assist platforms in formulating judicious resource allocation and operational strategies. Thirdly, in the computation of word sentiment values and text sentiment values, this study employs a suite of algorithms tailored to the specific textual characteristics inherent in freelance platforms, including the Naive Bayes algorithm and the TF-IDF algorithm. These algorithms are pivotal in refining the accuracy of feature value calculations, thereby imparting valuable insights for addressing analogous challenges.

5.2. Future outlook

To further optimize the performance of this method, there are three intriguing tasks for investigation in our future work:

(1) In constructing the Word Sentiment Value Calculation Model, drawing upon text data from a more comprehensive range of freelance platforms is imperative.

The greater the abundance and precision of the text data derived from freelance platforms for the computation of word sentiment values, the more closely aligned the calculated sentiment values of words will be to their true counterparts. As a result, there is a simultaneous enhancement in the predictive accuracy of the method when applied in practical scenarios. Consequently, our forthcoming endeavors in developing the Word Sentiment Value Calculation Model involve strategically utilizing text data sourced from a broader spectrum of freelance platforms.

(2) Extraction of Multidimensional Textual Features.

Currently, extracting features from textual data is limited to unidimensional attributes, with a predominant focus on sentiment values. However, the incorporation of multiple dimensions holds the promise of enhancing the interpretability of the model for the predicted outcomes. As a result, our forthcoming endeavors involve the comprehensive extraction of features from textual data across diverse dimensions, with the overarching goal of substantively improving the model's predictive accuracy.

(3) Exploring Textual Contextual Associations

The presence of sentiment words is intricately interconnected and not mutually independent. To optimize the selection of sentiment words, as well as the computation of word sentiment values and the calculation of text sentiment values, there are prospective plans to investigate a feather extraction methodology for sentiment analysis based on contextual factors, enhancing the precision of text sentiment value calculations.

References

[1] Roy G and Shrivastava A K 2020 Future of gig economy: opportunities and challenges. *IMI. Konnect* 9(1) 14–27.
[2] Ludwig S, Herhausen D, Grewal D, Bove L, Benoit S, De Ruyter K and Urwin P 2022 Communication in the gig economy: Buying and selling in online freelance marketplaces. *J. Marketing* 86(4) 14–161.

[3] El Barachi M, AlKhatib M, Mathew S and Oroumchian F 2021 A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. *J. Clean. Prod* 312 127820.
[4] Jain P K, Pamula R and Srivastava G 2021 A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Comput. Sci. Rev* 41 100413.
[5] Zhang J, Zhang A, Liu D and Bian Y 2021 Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews. *Knowl-Based. Syst* 228 107259.
[6] Liu J, Zhou Y, Jiang X and Zhang W 2020 Consumers' satisfaction factors mining and sentiment analysis of B2C online pharmacy reviews. *BMC. Med. Inform. Decis* 20 1–13.
[7] Graessley S, Horak J, Kovacova M, Valaskova K and Poliak M 2019 Consumer attitudes and behaviors in the technology-driven sharing economy: Motivations for participating in collaborative consumption. *J. Self-Gov. Manag. Bus* 7(1) 25–30.
[8] Taboada M 2016 Sentiment analysis: An overview from linguistics. *Annu. Rev. Linguist* 2 325–47.
[9] Ahmed M, Chen Q and Li Z 2020 Constructing domain-dependent sentiment dictionary for sentiment analysis. *Neural. Comput. Appl* 32 14719–32.
[10] Frankel R, Jennings J and Lee J 2022 Disclosure sentiment: Machine learning vs. dictionary methods. *Manage. Sci* 68(7) 5514–32.
[11] Benarafa H, Benkhalifa M and Akhloufi M 2023 WordNet Semantic Relations Based Enhancement of KNN Model for Implicit Aspect Identification in Sentiment Analysis. *Int. J. Comput. Int. Sys* 16(1) 3.
[12] Madani Y, Erritali M, Bengourram J and Sailhan F 2020 A hybrid multilingual fuzzy-based approach to the sentiment analysis problem using SentiWordNet. *Int. J. Uncertain. Fuzz* 28(03) 361–90.
[13] Karyawati A E, Utomo P A and Wibawa I G A 2022 Comparison of SVM and LIWC for Sentiment Analysis of SARA. *Ind. J. Comput. Cy. S* 16(1) 45–54.
[14] Wilksch M and Abramova O 2023 PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets. *Int. J. Inf. Ma. Data. Insig* 3(1) 100171.
[15] Xia L 2023 Chinese Financial Comments Sentiment Detection Based on the Bert-TCN Model Based on HowNet Disambiguation. *Proc. Int. Conf. on Digital Economy and Computer Application (DECA 2023)* (Shanghai: Shanghai-China/ Atlantis Press) pp 153–168
[16] Li H, Ma Y, Ma Z and Zhu H 2021 Weibo text sentiment analysis based on bert and deep learning. *Appl. Sci* 11(22) 10774.
[17] Okango E and Mwambi H 2022 Dictionary based global Twitter sentiment analysis of coronavirus (Covid-19) effects and response. *Ann. Data. Sci* 9(1) 175–86.
[18] Al Dujaili M J, Ebrahimi-Moghadam A and Fatlawi A 2021 Speech emotion recognition based on SVM and KNN classifications fusion. *Int. J. Electr. Comput* 11(2) 1259.
[19] Zhao Q 2021 Social emotion classification of Japanese text information based on SVM and KNN. *J. Amb. Intel. Hum. Comp* 1–12.
[20] Wang Z, Jiao R and Jiang H 2020 Emotion recognition using WT-SVM in human-computer interaction. *J. New. Media* 2(3) 121.
[21] Li Z, Li R and Jin G 2020 Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary. *Ieee. Access* 8 75073–84.
[22] Aslan M 2022 CNN based efficient approach for emotion recognition. *J. King. Saud. Univ-Com* 34(9) 7335–46.
[23] Liu Z X, Zhang D G, Luo G Z, Lian M and Liu B 2020 A new method of emotional analysis based on CNN–BiLSTM hybrid neural network. *Cluster. Comput* 23 2901–13.
[24] Yang M, Xu J, Luo K and Zhang Y 2021 Sentiment analysis of Chinese text based on Elmo-RNN model. *J. Physics: Conference Series* vol 1748 (Shenyang: Liaoning-China/ IOP Publishing) p 022033

DOI: 10.54254/2753-8818/34/20241167

[25] Zhang L 2021 Research on case reasoning method based on TF-IDF. *Int. J. Syst. Assur. Eng* 12 608–15.

[26] Feldman J, Zhang D J, Liu X and Zhang N 2022 Customer choice models vs. machine learning: Finding optimal product displays on Alibaba. *Oper. Res* 70(1) 309–28.

[27] Yu Q, Yang S, Zhang Z, Zhang Y L, Hu B, Liu Z, Huang K, Zhong X, Zhou J and Fang Y 2021 A graph attention network model for gmv forecast on online shopping festival. *Proc. Int. Conf. on Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data Proc* vol 12858 (Guangzhou:Guangzhou-China/ Cham: Springer International Publishing) pp 134–9

[28] Pan H and Zhou H 2020 Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce. *Electron. Commer. Res* 20 297–320.

[29] Ma S and Fildes R 2020 Forecasting third-party mobile payments with implications for customer flow prediction. *Int. J. Forecasting* 36(3) 739–60.

[30] Ding J, Chen Z, Xiaolong L and Lai B 2020 Sales forecasting based on catboost. *Proc. Int. Conf. on information technology and computer application (ITCA)* (Guangzhou: Guangdong-China/ IEEE) pp 636–39