# Multi-objective Decision Making Model for Stock Price Prediction Using Multi-source Heterogeneous Data Fusion

**Noura Metawa[1,*], Maha Mutawea[2]**

[1] College of Business administration, University of Sharjah, UAE

[2] Delta University for Science and Technology, Gamasa, Egypt

Emails: nmetawa@sharjah.ac.ae ; maha.mutawea@delta.gove.edu.eg

**Abstract**

Stock exchanges are developed as an essential component of economies, as they can promote financial and capital gain. The stock market is network of economic connections where share is bought and sold. Stock Market Prediction (SMP) is quite useful to investors. An effective forecast of stock prices is offer shareholders with suitable help in making appropriate decisions regarding if sell or purchase shares. The employ of Machine Learning (ML) and Sentiment Analysis (SA) on data in microblogging sites are developed as a famous approach to SMP. However, the heterogenous data fusion in stock market field is a big challenge. This paper introduces an effective Cat Swarm Optimization with Machine Learning Enabled Microblogging Sentiment Analysis for Stock Price Prediction technique. The presented model investigates the social media sentiments to foresee SPP. Firstly, the proposed model executes data pre-processing and Glove word embedding approach. Next, the weighted extreme learning machine approach was utilized for the classification of sentiments for SPP. Lastly, the CSO system was exploited for optimal adjustment of the parameters related to the WELM model. The experimental validation of the proposed approach was executed using microblogging data. The results show that the proposed method outperforms the previous studies.

**Keywords:** Sentiment analysis; Microblogging; Stock price prediction; Heterogeneous Data Fusion; Machine learning; forecasting model

## 1. Introduction

Stock price prediction (SPP) has been a functioning area of examination, in any case, it has forever been quite difficult for quantitative experts and scientists. As indicated by efficient market hypothesis (EMH) [1], foreseeing the market with 100 percent accuracy is inconceivable. However, this has not halted merchant's or scientists' endeavors in further developing methods to beat the random walk example of SPs. Various investigations show that stock market (SM) prices don't undergo a random walk and can without a doubt be anticipated somewhat. Recognizing the stock examples and outer variables is viewed as the key to working on the SPP [2]. The efficient market hypothesis places that SPs are an element of data and reasonable assumptions so SPs (Stock price) and financial markets are sentiment driven [3]. Conversely, a specific piece of SP is driven by hypotheses about how the organization would do from now on. With the accessibility of huge corpus of suppositions through virtual entertainment and the progress of machine learning (ML) techniques for enormous scope text mining, utilizing the clients' perspectives to further develop SPP has turned into a functioning subject of examination [4]. Many related explorers focus closer on the microblog text analysis and the prediction technique for SM, while the information filtering doesn't draw a lot of consideration. As the crude



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

information is bounteous and uproarious, stacking the information will be all an exercise in futility and even upset the result [5].

SPP is famous for the ascent of web finance. The EMH (Efficient Market Hypothesis) proposes that SM completely mirrors generally accessible data, and follows a random example [6]. This hypothesis has been generally acknowledged by financial experts over the previous years, yet presently individuals accept that we can portray a tip of the SM. Numerous strategies from different fields have shown promising outcomes around here. With the advantages of adequate web-based information and AI strategies [7], a few specialists apply NLP (natural language processing) methods in news, message sheets, and tweets to prepare a specific example to cover the SM development. These works utilize various techniques and accomplish remarkable outcomes. However, there actually need a few endeavors to work on the exhibition [8]. This work expands the exploration by adding the microblog filter and user group model, which are demonstrated viable to foresee the SM development. Enlivened by their work, sentiment starts to assume a significant part in SPP research. The financial backer's sentiment hypothesis likewise states that the merchant's venture conduct will be affected by sentiment.

StockTwits is a micro-blogging stage like Twitter. It is turning out to be progressively famous for clients to share their sentiments related to stocks [9]. Like Twitter, because as far as possible on StockTwits, feelings communicated are relevant and concise, moderately without any trace of common commotion remembered for some other news/blogging/virtual entertainment sites. StockTwits is intensely based on financial information and its clients are generally having knowledge regarding the financial markets [10]. Subsequently, it tends to be estimated that conclusions communicated through StockTwits are bound to impact the SMs and be more appropriate in SM prediction.

This paper introduces an effective cat Swarm Optimization with Machine Learning Enabled Microblogging Sentiment Analysis for Stock Price Prediction (CSOML-MSASPP) technique. The presented CSOML-MSASPP model investigates the social media sentiments to foresee SPP. Firstly, the CSOML-MSASPP model executes data pre-processing and Glove word embedding approach. Next, the weighted extreme learning machine (WELM) system was utilized for the classification of sentiments for SPP. Lastly, the CSO technique was exploited for optimal adjustment of the parameters related to the WELM model. The experimental validation of the CSOML-MSASPP approach was implemented using microblogging data.

## 2. Related works

In [11], we projected a method for foreseeing stock development using SA on Twitter and StockTwits information. Stock development and sentiment information is utilized to assess this methodology and approve it on Microsoft stock. They accumulated tweets from Twitter and StockTwits and financial information from Finance Yahoo. The principal curiosity of this work is that it incorporates numerous ML and SA strategies, stressing the recovery of additional features from online entertainment for further developing stock prediction exactness. Keramatfar et al. [12] paper use charts to address microblog posts and their different connections, like client, kinship, hashtag, sentimental similitude, printed closeness, and normal companions. It then, at that point, utilizes chart neural networks to perform setting mindful sentiment analysis (SA). For making utilization of the knowledge contained in numerous charts, we suggest a stacking method that at the same time utilizes different diagram types. The discoveries exhibit the pertinence of humanistic hypotheses in the analysis of virtual entertainment.

Basiri et al. [13] suggested a new technique based on the grouping of four deep learning and one traditional ML technique for SA of Covid related tweets from 8 nations. Likewise, we broke down Covid related look by utilizing Google Trends to more readily comprehend the adjustment of the sentiment design in various environments. Our discoveries uncover that the Covid pulled in the consideration of individuals from various nations at various times in shifting forces. Likewise, the sentiment in its tweets is related to the information and occasions that happened in their nations including the quantity of recently tainted cases, amount of recoveries, and passings. Yıldırım et al. [14] expect to perform SA of a financial microblog, to be specific, StockTwits. We completed the analysis on named messages of twelve stocks for a time of five months going from May 2019 to September 2019 utilizing different DL approaches. We thought about the presentation of the DL classifiers with conventional ML drawing near. Long Short-Term Memory (LSTM) model and its varieties, for example, bidirectional LSTM and bi-directional LSTM with dropout beat different classifiers.



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

However, utilization of dropout system didn't work on the presentation of the model yet there was a reduction in predisposition and difference.

Liu et al. [15] propose another technique to compute the correlation by utilizing the endeavor knowledge diagram implanting that efficiently thinks about different sorts of connections between recorded stocks. Additionally, we utilize Gated Recurrent Unit (GRU) model to join the associated stocks' news sentiment, the central stock's news sentiment, and the central quantitative features of stocks to anticipate the central SP development. In [16], we propose a technique for breaking down and anticipating SPs of organizations utilizing the previously mentioned algorithms as a group. Datasets from India's National Stock Exchange (NSE) containing fundamental market price data are preprocessed to incorporate notable driving technical indicators as features. Include determination, which ranks features based on their level of effect on the last closing price has been integrated to lessen the size of the preparation dataset. Furthermore, we assess the viability of considering the general assessment of an organization by utilizing SA. Utilizing a prepared Word2Vec model, organization explicit hash-labeled posts from Twitter are named positive or negative. In [17], we will coordinate a few factors that can influence the SPs by coordinating SA with important printed features with relevant slacks with the mean to develop more solid and reasonable sentiment indication.

## 3. The Proposed Model

In this study, a new CSOML-MSASPP approach was developed to investigate the microblogging sentiments for SPP. The presented CSOML-MSASPP model investigates the social media sentiments to foresee SPP. Firstly, the CSOML-MSASPP model executes data pre-processing and Glove word embedding approach. Next, the WELM system was utilized for the classification of sentiments for SPP. Finally, the CSO technique was exploited for optimal adjustment of the parameters related to the WELM model. Fig. 1 demonstrates the block diagram of CSOML-MSASPP approach.

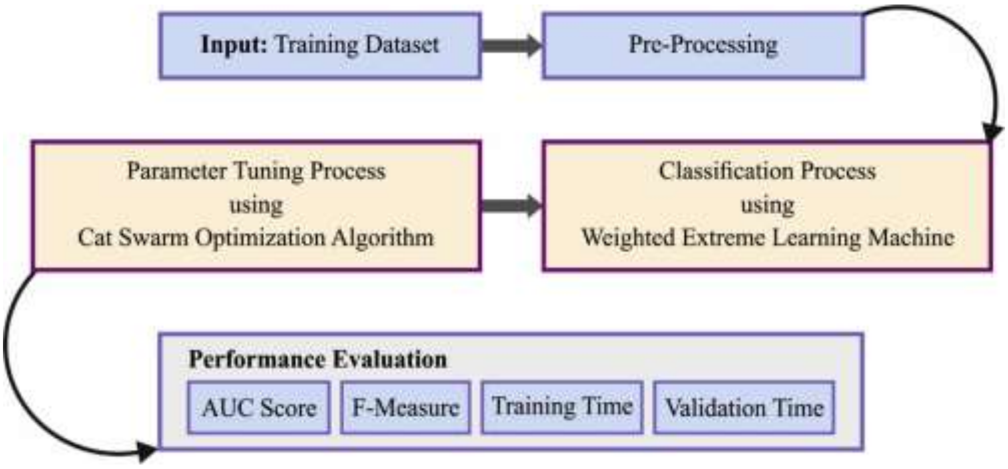


Figure 1: Block diagram of Multi-source Heterogeneous Data Fusion approach

### 3.1 Data Pre-processing and Word Embedding

In the first stage, the CSOML-MSASPP model executes data pre-processing and Glove word embedding approach [18]. As Word2vec doesn't consider global context, hence Glove neural word embedding came into picture. Glove embedding uses the similar intuition behindhand distribution embedding of co-occurrence matrixes, the only possible distinction is that it uses a neural network to decompose occurrence matrixes into compact word vectors. Glove word vector has demonstrated more efficient outcome when compared to Word2vec in word analogy task as Glove more meaning is added to neural word embedding by considering the relationships amongst word pair to word pain. Additionally, Glove allocates lesser weight to increasingly frequent word pairs including "a", "the", and so on. But the model depends on co-occurring matrixes, therefore Glove needs a large quantity of storage. As well, changing hyperparameter is highly connected with the co-occurring matrixes, one needs to recreate the whole matrixes again which will use up a considerable amount of time.



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

### 3.2 WELM based Classification

Next to data pre-processing and word embedding, the WELM system was utilized for the classification of sentiments for SPP. An ELM is exploited for the classifier of balanced data, whereas WELM is exploited for the classifier of imbalanced data [19]. The trained data has $N$ different samples, $(x_i, z_i), i = 1,2, \dots, N$. A neural network (NN) and a single hidden layer (HL) using $L$ HL nodes are formulated by:

$$\sum_{i=1}^{L} \beta_i \cdot 1(w_i \cdot x_j + b_i) = z_j, j = 1, \dots, N \quad (1)$$

In Eq. (1), $w_i$ indicates a single HL input weight,$l()$ denotes the activation function, $\beta_i$ represent the output weight, and $b_i$ signifies single HL bias that is defined in the following.

$$S\beta = T \quad (2)$$

In Eq. (2), $S$ indicates the resulting matrixes of single HL.

$$S(w_1, \dots, w_L, b_1, \dots, b_L, x_1, \dots, x_N) = \begin{pmatrix} l(w\_1 \cdot x\_1 + b_1 & \cdots & l(w_L x_1 + b_L) \\ \vdots & \ddots & \vdots \\ l(w_1 \cdot x_N + b_1) & \cdots & l(w_L \cdot x_N + b_L) \end{pmatrix}_{N \times L} \quad (3)$$

According to the Karush–Kuhn-Tucker model, the Lagrangian factor is recognized in varying the trainable ELM as various difficulties are managed. The $\beta$ resulting weight is estimated as follows:

$$\beta = S^T \left(\frac{1}{C} + SS^T\right)^{-1} T \quad (4)$$

Here, $C$ denotes the regularization coefficients. Thus, the output function of ELM classifier is formulated by:

$$F(x) = s(x)S^T(\frac{1}{C} + SS^T)^{-1}T = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{1}{C} + \chi\right)^{-1} T \quad (5)$$

In Eq. (5), $\chi$ denotes the kernel matrix that is estimated by.

$$\chi = SS^T = s(x_i)s(x_j) = K(x_i, x_j) \quad (6)$$

It can be noticeable that the HL feature maps $s(x)$ represent the individuality in classifier result of ELM. The classifier result was compared to the kernel function, $K(x, y)$. $K(x, y)$ consider an internal creation procedure:

$$K(x, y) = \exp\left(-\gamma \|x - y\|^2\right) \quad (7)$$

As a result, the kernel ELM (KELM) classifier efficiency was distributed into two variables such as the $\gamma$ kernel function variable, and C penalty variable. The actual benefits of ELM are obtained in WELM by maintaining the weight for different samples for handling imbalanced classifier problem. It is determined by the subsequent expression:

$$F(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^{\mathrm{T}} \left(\frac{1}{C} + W\chi\right)^{-1} WT, \quad (8)$$

$$W = diag\,(w_{ii}), i = 1,2, \dots, N \quad (9)$$

Now, $W$ symbolizes the weighted matrixes. WELM has two weightage modules as follows:



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

$$w_{ii} = \frac{1}{\#(z_i)}, \tag{10}$$

$$w_{ii} = \begin{cases} \frac{0.618}{\#(z_i)}, if z_i > \bar{z} \\ \frac{1}{\#(z_i)}, otherwise \end{cases} \tag{11}$$

Here, $\#(z_i)$ represents the instance count, goes to class $z_i$, $i = 1, \dots, m$. $m$ designates the class count. $\underline{z}$ means the average of wide-ranging examples. Fig. 2 displays the infrastructure of ELM model.

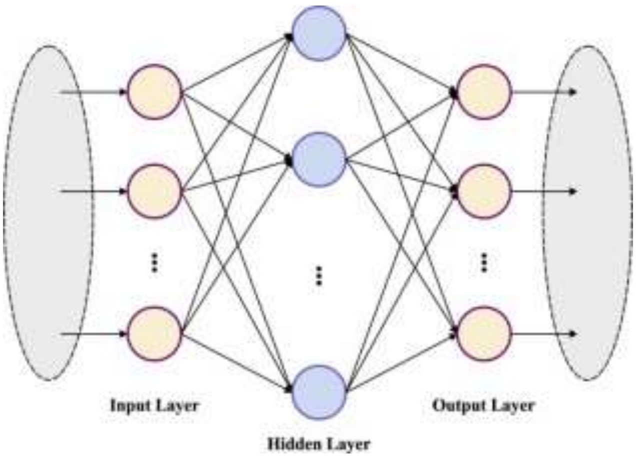


Figure 2: Structure of ELM

### 3.3 Parameter Optimization using CSO Algorithm

In the final stage, the CSO algorithm is exploited for optimal adjustment of the parameters related to the WELM model. The CSO method is depending on two most important characteristics of the cats namely resting and hunting [20]. In the initial phase, the cat spends its time resting, where they stay alert and moves slowly to alternative location. On the other hand, once a target was identified, the cat seizures the target. Therefore, numerical modelling is calculated to decide on challenging optimization problems and is entitled afterward the CSO process. In this system, tracing and seeking methods are determined to define the cat's behavior. The working procedure of these modes is defined in the following.

**Seeking Mode**

This technique determines the resting mode. Now, they move towards various positions in the searching region; notwithstanding this, they stay alert. It is understood as a local search for the resolution. The subsequent symbolization is exploited in these models.

Seeking Memory Pool (SMP): it determines the copy count of the cat that has been simulated.

Seeking Range of selected dimension (SRD): it denotes the modification amongst older and newer dimensions of the cat selected for mutation.

Count of Dimension to Change (CDC): it denotes the dimension count of a cat position, practiced for mutation.

Define the count of copies (T) of $i$-$th$ cats. As per CDC variable, add or subtract the SRD values at random from the present place of cat, and replaces the older value for all the copies.

Evaluate the fitness of all the copies

Choose the optimum solution candidate and deployed it in the location of $i$-$th$ cats.



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

**Tracing Mode**

The tracing mode reflects the hunting mode of the cat. As soon as the cat hunts the prey, the location and velocity of the cat are upgraded. Therefore, a larger variance occurs among older and newer positions of the cat. The position $(X_j^d)$ and velocity $(V_j^d)$ of $j\text{-}th$ cat in $d$—dimension vector is defined as $X_j^d = \{X_j^1, X_j^2, \dots \dots . X_j^D\}; V_j^d = \{V_j^1, V_j^2, \dots \dots . V_j^D\}$; where $d = 1,2, \dots$ The optimum placed of the cat is epitomized by $X_{best}^d = \{X_{best}^1, X_{best}^2, \dots \dots . X_{best}^D\}$. The location and velocity of $j\text{-}th$ cat are evaluated by the subsequent formula:

$$V_{jnew}^d = w * V_j^d + c * r * \left(X_{jbest}^d - X_j^d\right) \tag{12}$$

In Eq. (12), $w$ indicates a weight factor among zero and one, $V_{jnew}^d$ symbolizes the upgraded velocity of $j\text{-}th$ cat in $d\text{-}th$ dimensions, $c$ denotes a user-defined constant, $V_j^d$ means the older velocity of $j\text{-}th$ cats, $X_{jbest}^d$ indicates the optimum position obtained using $j\text{-}th$ cat in $d\text{-}th$ dimensions, $r$ denotes an arbitrary value among zero and one, and $X_j^d$ indicates the present position of $j\text{-}th$ cat in $d\text{-}th$ dimensions, whereas $d = 1,2, \dots D$.

$$X_{jnew}^d = X_j^d + V_j^d \tag{13}$$

In Eq. (13), $X_j^d$ denotes the present location of $j\text{-}th$ cat in $d$-dimension, $V_j^d$ indicates the velocity of $j\text{-}th$ cat, $and\ X_{jnew}^d$ characterizes the upgraded position of $j\text{-}th$ cat in $d$-dimensional. A mixture ratio (MR) was exploited for combining trace and seek modes. An MR is designed for describing the cat count in trace and seek modes.

## 4. Performance Validation

The experimental validation of the CSOML-MSASPP model is performed using two datasets. Table 1 and Fig. 3 showcases the simulation results of the CSOML-MSASPP model interms of $F_{score}$ and $AUC_{score}$ on dataset-1 [11]. The results implied that the CSOML-MSASPP model has shown enhanced results with maximum values of $F_{score}$ and $AUC_{score}$. With respect to $F_{score}$, the CSOML-MSASPP model has gained increased $F_{score}$ of 77.32% whereas the KNN, SVM, LR, NB, DT, RF, and MLP models have demonstrated reduced $F_{score}$ of 53.74%, 69.07%, 60.68%, 67.82%, 62.24%, 62.74%, and 57.27% respectively. Also, in relation to $AUC_{score}$, the CSOML-MSASPP method has attained maximum $AUC_{score}$ of 71.23% whereas the KNN, SVM, LR, NB, DT, RF, and MLP methods have illustrated reduced $AUC_{score}$ of 40.43%, 54.64%, 44.62%, 52.34%, 50.78%, 50.36%, and 45.84% correspondingly.

Table 1: Result analysis of CSOML-MSASPP approach with existing methods under dataset-1

| Dataset-1 | | |
|---|---|---|
| **Model** | **F-Measure** | **AUC Score** |
| CSOML-MSASPP | 77.32 | 71.23 |
| KNN | 53.74 | 40.24 |
| SVM | 69.07 | 54.64 |
| LR | 60.68 | 44.62 |
| NB | 67.82 | 52.34 |
| DT | 62.24 | 50.78 |
| RF | 62.74 | 50.36 |
| MLP | 57.27 | 45.84 |



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

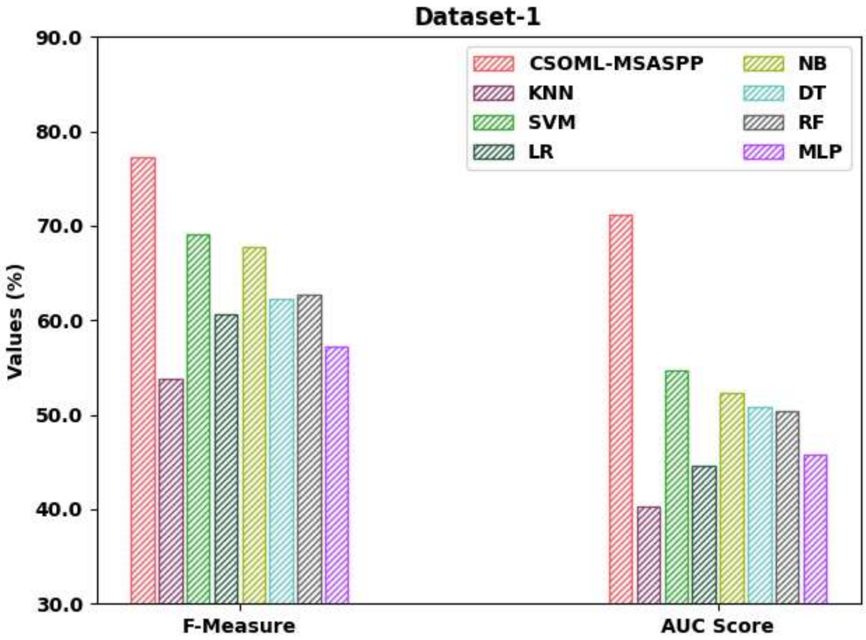


Figure 3: Result analysis of CSOML-MSASPP approach under dataset-1

Table 2 and Fig. 4 exhibits the simulation results of the CSOML-MSASPP technique interms of $F_{score}$ and $AUC_{score}$ on dataset-2. The results denoted the CSOML-MSASPP method has shown enhanced results with maximal values of $F_{score}$ and $AUC_{score}$. With regard to $F_{score}$ the CSOML-MSASPP algorithm has acquired increased $F_{score}$ of 72.92% whereas the KNN, SVM, LR, NB, DT, RF, and MLP methodologies have demonstrated reduced $F_{score}$ of 45.77%, 67.20%, 68.35%, 67.36%, 53.09%, 58.05%, and 66.19% correspondingly. Along with that, with respect to $AUC_{score}$, the CSOML-MSASPP approach has reached increased $AUC_{score}$ of 72.28% whereas the KNN, SVM, LR, NB, DT, RF, and MLP methods have demonstrated reduced $AUC_{score}$ of 39.67%, 54.62%, 54.43%, 50.69%, 45.87%, 46.26%, and 51.19% correspondingly.

Table 2: Result analysis of CSOML-MSASPP approach with existing methods under dataset-2

| **Dataset-2** | | |
|---|---|---|
| **Model** | **F-Measure** | **AUC Score** |
| CSOML-MSASPP | 72.92 | 72.28 |
| KNN | 45.77 | 39.67 |
| SVM | 67.20 | 54.62 |
| LR | 68.35 | 54.43 |
| NB | 67.36 | 50.69 |
| DT | 53.09 | 45.87 |
| RF | 58.05 | 46.26 |
| MLP | 66.19 | 51.19 |



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

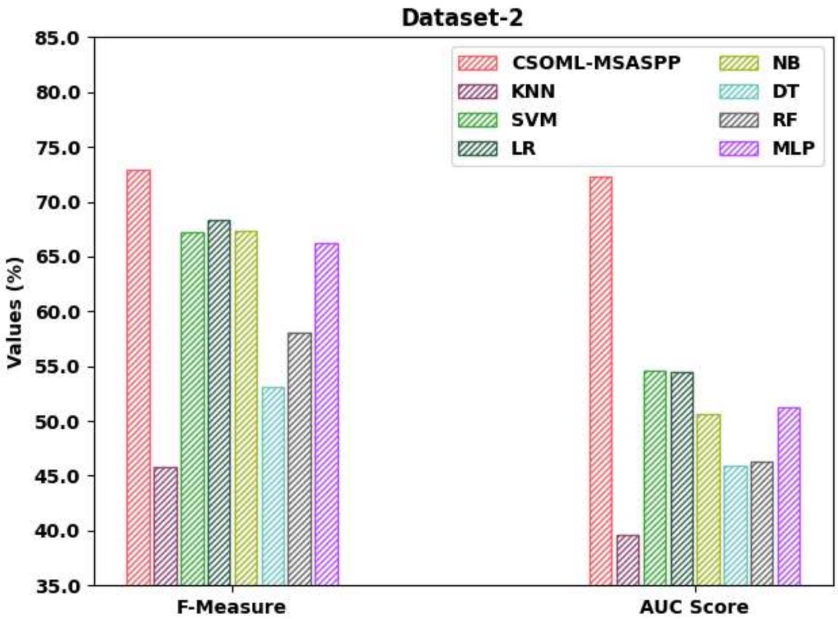


Figure 4: Result analysis of CSOML-MSASPP approach under dataset-1

Table 3 and Fig. 5 show the simulation results of the CSOML-MSASPP technique in terms of $F_{score}$ and $AUC_{score}$ on dataset-3. The results denoted that the CSOML-MSASPP model has shown enhanced results with maximum values of $F_{score}$ and $AUC_{score}$. In relation with $F_{score}$, the CSOML-MSASPP algorithm has acquired increased $F_{score}$ of 73.12% whereas the KNN, SVM, LR, NB, DT, RF, and MLP approaches have exhibited reduced $F_{score}$ of 45.50%, 67.48%, 69.18%, 67.35%, 53.53%, 58.18%, and 67.82% correspondingly. Furthermore, in connection with $AUC_{score}$, the CSOML-MSASPP technique has gained maximum $AUC_{score}$ of 68.54% whereas the KNN, SVM, LR, NB, DT, RF, and MLP methodologies have demonstrated minimal $AUC_{score}$ of 39.02%, 54.45%, 56.08%, 51.14%, 45.81%, 44.03%, and 48.99% correspondingly.

Table 3: Result analysis of CSOML-MSASPP approach with existing methods under dataset-3

| Dataset-3 | | |
|---|---|---|
| **Model** | **F-Measure** | **AUC Score** |
| CSOML-MSASPP | 73.12 | 68.54 |
| KNN | 45.50 | 39.02 |
| SVM | 67.48 | 54.45 |
| LR | 69.18 | 56.08 |
| NB | 67.35 | 51.14 |
| DT | 53.53 | 45.81 |
| RF | 58.18 | 44.03 |
| MLP | 67.82 | 48.99 |

Table 4 and Fig. 6 illustrate the simulation results of the CSOML-MSASPP approach in relation to of $F_{score}$ and $AUC_{score}$ on dataset-4. The results implied that the CSOML-MSASPP method has shown enhanced results with maximal values of $F_{score}$ and $AUC_{score}$.



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

Table 4: Result analysis of CSOML-MSASPP approach with existing methods under dataset-4

| Dataset-4 | | |
|---|---|---|
| **Model** | **F-Measure** | **AUC Score** |
| CSOML-MSASPP | 79.33 | 75.28 |
| KNN | 71.34 | 69.53 |
| SVM | 76.29 | 44.98 |
| LR | 56.14 | 51.19 |
| NB | 71.36 | 49.17 |
| DT | 71.78 | 67.75 |
| RF | 70.51 | 65.81 |
| MLP | 75.96 | 49.17 |

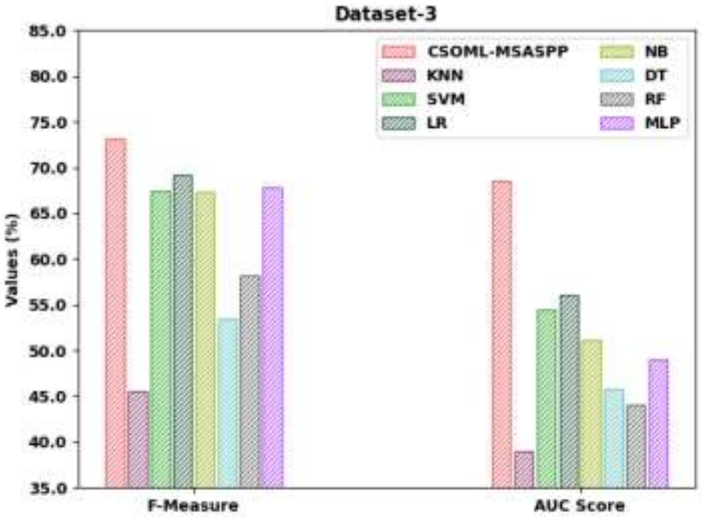


Figure 5: Result analysis of CSOML-MSASPP approach under dataset-3

In line with $F_{score}$, the CSOML-MSASPP approach has obtained increased $F_{score}$ of 79.33% whereas the KNN, SVM, LR, NB, DT, RF, and MLP methods have demonstrated reduced $F_{score}$ of 71.34%, 76.29%, 56.14%, 71.36%, 71.78%, 70.51%, and 75.96% correspondingly. Along with that, with regard to $AUC_{score}$, the CSOML-MSASPP technique has reached increased $AUC_{score}$ of 75.28% whereas the KNN, SVM, LR, NB, DT, RF, and MLP models have established reduced $AUC_{score}$ of 69.53%, 44.53%, 51.19%, 49.17%, 67.75%, 65.81%, and 49.17% correspondingly.

Table 5 and Fig. 7 highlight the TRT and VT outcomes of the CSOML-MSASPP model with existing models. With respect to TRT, the CSOML-MSASPP model has shown enhanced performance with lower TRT of 3.54s whereas the KNN, SVM, LR, NB, DT, RF, and MLP models have demonstrated reduced outcomes with TRT of 4.85s, 4.04s, 2.88s, 6.57s, 10.35s, 10.70s, and 21.65s respectively.

Table 5: TRT and VT analysis of CSOML-MSASPP system with existing approaches

| Methods | Training Time (s) | Validation Time (s) |
|---|---|---|
| CSOML-MSASPP | 3.54 | 1.19 |
| KNN | 4.85 | 1.71 |
| SVM | 4.04 | 1.36 |
| LR | 2.88 | 0.98 |
| NB | 6.57 | 2.67 |
| DT | 10.35 | 3.45 |
| RF | 10.70 | 3.56 |
| MLP | 21.65 | 9.59 |



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

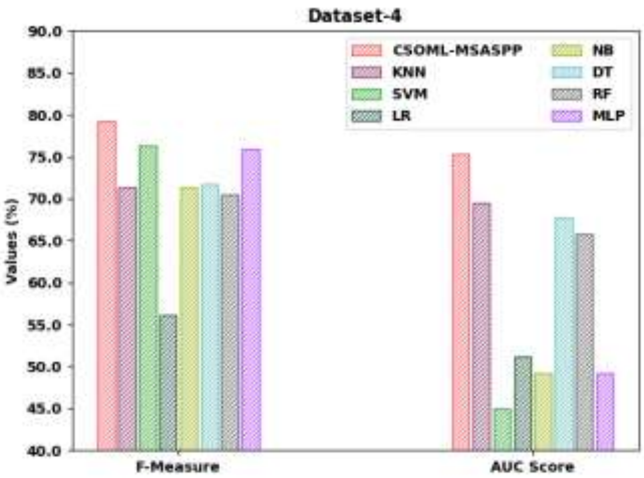


Figure 6: Result analysis of CSOML-MSASPP approach under dataset-4

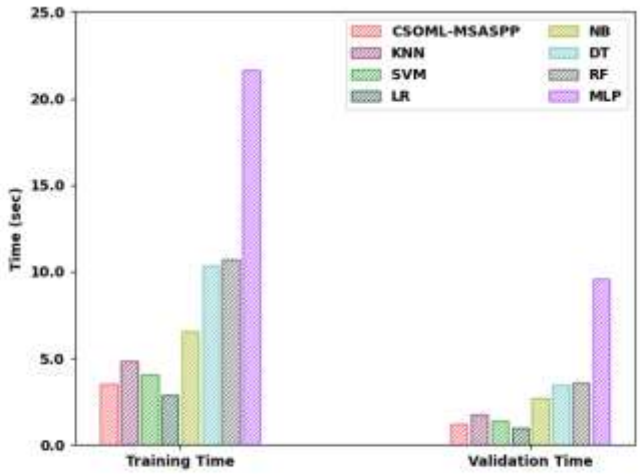


Figure 7: TRT and VT analysis of CSOML-MSASPP system with existing approaches

Also, with regard to VT, the CSOML-MSASPP technique has shown enhanced performance with lower VT of 1.19s whereas the KNN, SVM, LR, NB, DT, RF, and MLP methodologies have demonstrated reduced outcomes with VT of 1.71s, 1.36s, 0.98s, 2.67s, 3.45s, 3.56s, and 9.59s correspondingly.

## 5. Conclusion

In this study, a new model was developed to investigate the microblogging sentiments for SPP based on multi-source heterogeneous data fusion. The presented model investigates the social media sentiments to foresee SPP. Firstly, the model executes data pre-processing and Glove word embedding approach. Next, the WELM approach was utilized for the classification of sentiments for SPP. Lastly, the CSO system was exploited for optimal adjustment of the parameters related to the WELM model. The experimental validation of the proposed approach was executed utilizing microblogging data. The comparative study reported the enhanced outcomes of the proposed method over existing methodologies. As a part of future scope, the CSOML-MSASPP model can be applied to forecast customer churns in real time.

## 5. References

[1] Chun, J., Ahn, J., Kim, Y. and Lee, S., 2021. Using deep learning to develop a stock price prediction model based on individual investor emotions. Journal of Behavioral Finance, 22(4), pp.480-489.



Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022

[2] Hajiali, M., 2020. Big data and sentiment analysis: A comprehensive and systematic literature review. Concurrency and Computation: Practice and Experience, 32(14), p.e5671.

[3] Gupta, I., Madan, T.K., Singh, S. and Singh, A.K., 2022. HiSA-SMFM: Historical and Sentiment Analysis based Stock Market Forecasting Model. arXiv preprint arXiv:2203.08143.

[4] Daudert, T., 2021. Exploiting textual and relationship information for fine-grained financial sentiment analysis. Knowledge-Based Systems, 230, p.107389.

[5] Issam, A., Mounir, A.K. and El Mendili Saida, E.M.F., 2022. Financial sentiment analysis of tweets based on deep learning approach. Indonesian Journal of Electrical Engineering and Computer Science, 25(3), pp.1759-1770.

[6] Hossain, M.S., Rahman, M.F., Uddin, M.K. and Hossain, M.K., 2022. Customer sentiment analysis and prediction of halal restaurants using machine learning approaches. Journal of Islamic Marketing, (ahead-of-print).

[7] Nawaz, U., Ali, A., Raza, U.A. and Shehzadi, K., 2021. A Survey: Sentimental Analysis on Product Reviews Using (MLT) Machine Learning Techniques and Approaches. International Journal, 10(2).

[8] Kanakaraddi, S.G., Chikaraddi, A.K., Gull, K.C. and Hiremath, P.S., 2020, February. Comparison study of sentiment analysis of tweets using various machine learning algorithms. In 2020 International Conference on Inventive Computation Technologies (ICICT) (pp. 287-292). IEEE.

[9] Wadawadagi, R.S. and Pagi, V.B., 2020. Sentiment analysis on social media: Recent trends in machine learning. Handbook of Research on Emerging Trends and Applications of Machine Learning, pp.508-527.

[10] Malawana, M.V.D.H.P. and Rathnayaka, R.T., 2020, December. The Public Sentiment analysis within Big data Distributed system for Stock market prediction–A case study on Colombo Stock Exchange. In 2020 5th International Conference on Information Technology Research (ICITR) (pp. 1-6). IEEE.

[11] Koukaras, P., Nousi, C. and Tjortjis, C., 2022, May. Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning. In Telecom (Vol. 3, No. 2, pp. 358-378). MDPI.

[12] Keramatfar, A., Amirkhani, H. and Bidgoly, A.J., 2022. Modeling Tweet Dependencies with Graph Convolutional Networks for Sentiment Analysis. Cognitive Computation, pp.1-12

[13] Basiri, M.E., Nemati, S., Abdar, M., Asadi, S. and Acharrya, U.R., 2021. A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets. Knowledge-Based Systems, 228, p.107242

[14] Yıldırım, S., Jothimani, D., Kavaklioğlu, C. and Başar, A., 2019, December. Deep learning approaches for sentiment analysis on financial microblog dataset. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 5581-5584). IEEE

[15] Liu, J., Lu, Z. and Du, W., 2019, January. Combining enterprise knowledge graph and news sentiment analysis for stock price prediction. In Proceedings of the 52nd Hawaii International Conference on System Sciences

[16] Pasupulety, U., Anees, A.A., Anmol, S. and Mohan, B.R., 2019, June. Predicting stock prices using ensemble learning and sentiment analysis. In 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) (pp. 215-222). IEEE

[17] Bouktif, S., Fiaz, A. and Awad, M., 2019, October. Stock market movement prediction using disparate text features with machine learning. In 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS) (pp. 1-6). IEEE.

[18] Oota, S.R., Manwani, N. and Bapi, R.S., 2018, December. fMRI semantic category decoding using linguistic encoding of word embeddings. In International Conference on Neural Information Processing (pp. 3-15). Springer, Cham.

[19] Wang, Y., Wang, A., Ai, Q. and Sun, H., 2019. Ensemble based fuzzy weighted extreme learning machine for gene expression classification. Applied Intelligence, 49(3), pp.1161-1171.

[20] Huang, J., Asteris, P.G., Manafi Khajeh Pasha, S., Mohammed, A.S. and Hasanipanah, M., 2020. A new auto-tuning model for predicting the rock fragmentation: a cat swarm optimization algorithm. Engineering with Computers, pp.1-12.




Doi : https://doi.org/10.54216/FPA.090105
Received: May 18, 2022 Accepted: August 16, 2022