

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/387393551>

Innovative Hybrid Deep Learning Models for Financial Sentiment Analysis

Article in *Acadlore Transactions on AI and Machine Learning* · December 2024

DOI: 10.56578/ataiml030404

CITATION

1

READS

84

3 authors:



Ridwan Marqas
Firat University

37 PUBLICATIONS 355 CITATIONS

SEE PROFILE



Abdulazeez Mousa
Nawroz University

7 PUBLICATIONS 69 CITATIONS

SEE PROFILE



Fatih Özyurt
Firat University

61 PUBLICATIONS 1,969 CITATIONS

SEE PROFILE



Innovative Hybrid Deep Learning Models for Financial Sentiment Analysis

Ridwan B. Marqas^{1,2*}, Abdulazeez Mousa^{1,3}, Fatih Özyurt¹

¹ Department of Software Engineering, Firat University, 23119 Elazig, Turkey

² Department of Computer Science, Knowledge University, 44001 Erbil, Iraq

³ Department of Computer Science, Nawroz University, 42001 Duhok, Iraq

* Correspondence: Ridwan B. Marqas (ridwanmarqas@gmail.com)

Received: 10-25-2024

Revised: 12-13-2024

Accepted: 12-19-2024

Citation: R. B. Marqas, A. Mousa, and F. Özyurt, "Innovative hybrid deep learning models for financial sentiment analysis," *Acadlore Trans. Mach. Learn.*, vol. 3, no. 4, pp. 225–236, 2024. <https://doi.org/10.56578/ataiml030404>.



© 2024 by the authors. Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: This study explores hybrid deep learning architectures for the classification of financial sentiment, focusing on the integration of the Convolutional Neural Network (CNN) with the Support Vector Machine (SVM) and the Random Forest (RF). CNN, with its powerful feature extraction capabilities, was combined with SVM's ability to handle non-linear decision boundaries, while RF enhanced model generalization through ensemble learning. The proposed hybrid frameworks addressed two fundamental challenges in sentiment analysis: overfitting and class imbalance. These challenges were mitigated, resulting in improved model accuracy and reliability compared to standalone methods. Empirical evaluations demonstrated that the CNN-SVM model achieved competitive or superior validation accuracy and loss, indicating its suitability for precise financial sentiment classification. By enabling more accurate sentiment categorization, the model provides actionable insights for financial analysts and investors, thereby supporting better market assessment and investment decision-making. Future work is suggested to incorporate advanced techniques such as adversarial training and domain-specific pre-trained models to further enhance model performance.

Keywords: Finance; Deep learning; Convolutional Neural Network; Hybrid models; Sentiment analysis; Support Vector Machine; Random Forest

1 Introduction

The process of analyzing financial feelings with deep learning involves the application of deep learning methodologies to understand and assess the emotions and sentiments expressed by people and the general public on financial and economic topics. This research involves the use of financial and economic knowledge information such as financial reports, market information, and economic signals. Machine learning, particularly in the kind known as deep learning, was employed for the training of the computer models for the purpose of extractive processing of the stuff and financial data in particular. Such models include, for instance, the use of deep neural networks, which consist of complex structures designed to carry out highly detailed data computation and to recognize the existence of certain patterns or trends. The objective of this technology is to improve the decision-making process in investment by providing investors and financial analysts with better evaluations of the financial and economic environment. In addition, it can be used as a tool for market research to monitor the trends of financial markets as well as to predict the tendencies in the near future. Using deep learning approaches to analyze feelings about money, there is an opportunity for different parties, including persons, businesses, and organizations, to obtain a better understanding of financial markets and their behavior. Due to this improved perception, they can be in a better position to make better monetary decisions towards investments.

CNN integration with SVM and RF as a tool for financial sentiment classification represents a major leap forward in the use of deep learning applications in this area. SVM and RF are highly capable at classification, whereas CNN can extract in-depth textual data features. This study proposes models that combine the strengths of these approaches to tackle issues of sentiment analysis in finance.

The CNN-SVM hybrid model proposed in this study employs CNN to extract features and to discover semantic and syntactic patterns from financial texts such as articles, reports, and social media posts. SVM, being robust to

complex decision boundaries and its ability to curb overfitting while working in the multi-class scenario, classifies the extracted features. In addition, CNN-RF uses CNN to learn features and uses RF as an ensemble classifier to deal with high-dimensional feature space and noisy datasets to increase generalizability and robustness.

In contrast to regular deep learning methods that depend solely on neural net structure, these hybrid models prioritize interpretability and efficiency. The models leverage the strengths of traditional classifiers with CNN to resolve common problems, like class imbalance and overfitting, found in financial sentiment datasets. This integration increases both classification accuracy and reliance on prediction and gives a more extensive understanding of financial sentiments.

What these advances mean in practice, however, is profound. For investors and financial analysts, correctly labeling sentiments as positive, negative, or neutral helps them analyze market and economic trends in a better way. The proposed hybrid models have the ability to improve the decision-making processes, and their utilization in market research and investment strategies can substantially improve actionable insights extracted from many sources of financial data.

The selection between these two hybrid models depends on the nature and size of the dataset, the requirement for a specific sentiment analysis task, and computational capacities. Both of the hybrid models have the aim of increasing the accuracy and the reliability of sentiment analysis in financial contexts by combining outstanding deep learning feature extraction methods, namely CNN, with stable classification algorithms such as SVM or RF. In an attempt to compare the performance of both methodologies, researchers and analysts often conduct experimentation with one or the other and sometimes both.

2 Literature Review

Sentiment analysis has turned out to be one of the fundamental tools for financial analysis that helps stakeholders to extract actionable insights from unstructured big textual data sources such as news articles, social media, financial reports, etc. Following recent breakthroughs in machine learning and deep learning techniques, sentiment classification is becoming both more accurate and more efficient. This observation is especially notable as hybrid deep learning models have demonstrated an exceptional capability to learn and capture the most nuanced of all financial sentiments. This section synthesizes existing studies to offer a broad overview of the field.

Recently, CNN-based techniques have been shown to be an extremely powerful method for extracting high-level features from raw text data. Although Kim [1] demonstrated CNN's utility for sentence classification, it has not yet been applied to financial sentiment analysis. Following that, Zhang and Wallace [2] showed how sensitive CNN architectures are to different hyperparameters and the critical need to tune these hyperparameters. When these models are combined with cutting-edge classification algorithms, better sentiment analysis scores can be consistently obtained compared to traditional approaches. There are numerous studies that have applied deep learning to the problems of stock market predictions. Akita et al. [3] combined deep learning models to predict stock trends using numerical and textual data. Hu et al. [4] also proposed a framework with news data to predict trends, which involves textual analysis to examine market movements. Ding et al. [5, 6] further extended deep learning in financial contexts by introducing structured events with which they were able to predict stock price movements using structured financial data.

A lot of work has been done to explore how effective CNN in combination with SVM or RF would be for sentiment classification. Using topic-based sentiment analysis on Twitter data, Si et al. [7] revealed that hybrid approaches are indispensable where social media dynamics are concerned. Nguyen et al. [8] extended this work by applying lexicon-based approaches to sentiment analysis using StockTwits data, and found the shortcomings of only using lexicon-based approaches and hybrid approaches. The CNN-SVM model uses CNN for feature extraction and uses SVM for classification. Cortes and Vapnik [9] and Joachims [10] pointed out the good behavior of SVM for multi-classification problems and how it is immune to the order of some variables. The methods integrating CNN with RF have also been explored. For example, Breiman [11] discussed the ensemble capability of RF on handling large and noisy datasets.

Deep learning has recently brought forth novel techniques for performing sentiment analysis. For the case of financial time series forecasting, Widiputra et al. [12] used a CNN-LSTM framework as an appreciation of both the CNN for feature extraction and the LSTM for temporal modeling. Liu et al. [13] included adversarial training in their models and the authors were able to show how it can be beneficial in the quest to improve robustness by minimizing the effects that come with noisy or even adversarial inputs. These progress in turn stresses the usefulness of hybrid convolutional and recurrent structures in financial analysis, specifically in applications that demand spatial structuring as well as sequential features learning, Nelson et al. [14] employed the Long Short-Term Memory (LSTM) networks for stock movement prediction.

The development of FinancialBERT and more domain-specific models like it has been revolutionary for financial natural language processing (NLP). FinancialBERT, a pre-trained model for financial text mining, was introduced by Yang et al. [15] and is already achieving state-of-the-art results on sentiment classification tasks. All of this highlights that domain-specific adaptation contributes greatly to the performance of models.

Over the last couple of years, the use of the market variables with textual data has been a common feature in sentiment analysis. Wen et al. [16] showed that trends can be forecasted by means of high-order time series, providing means for integrating quantitative time series data with textual data for improved prediction. Some work that is similar to this has been done by Li et al. [17]. The elements of market fundamental and textual analysis have not been worked through as comprehensively as they are in this paper. By using this approach of integrating structured financial data indicators with unstructured textual sentiment, then the market behavior assessment can be analyzed from a more rounded perspective.

However, in spite of excellent progress made in the field of financial sentiment analysis, there are still many issues that deserve attention. As overfitting becomes an innate problem, as Nelson et al. [14] stated, regularization techniques and hyper parameter optimization have to be employed. In financial datasets, one of the issues is class imbalance, which can be addressed by weighted loss functions and advanced sampling techniques. According to Ding et al. [6], future research should address the interpretation of hybrid models so that they can enhance decision-making. Moreover, sentiment classification accuracy can be improved by integrating diverse data sources, such as the integration of structured financial reports with unstructured social media data. Feng et al. [18] reported emerging techniques, such as Generative Adversarial Network (GAN) and reinforcement learning, as interesting areas for future research.

The discussion has been further substantiated with the additional references which provide insights into distinct methodologies. The fundamentals of gradient-based learning lay the foundations for what modern CNN is built on today, as emphasized by LeCun et al. [19]. Peng and Jiang [20] and Liu et al. [13] revealed that loading word embedding with deep neural networks improves the accuracy of prediction in financial environments. This combination improves semantic and syntactic interpretability for real textual data to provide more accurate contextually-driven financial forecasts to the model.

Fine-grained opinion extraction is an important problem, and as a result, there have been great strides in the methods for aspect-based sentiment analysis (ABSA). Wankhade et al. [21] provided a survey of ABSA techniques for financial domains and found that they can be adapted. This landscape has been further enriched with the advent of multi-modal sentiment analysis. Pandey and Vishwakarma [22] conducted a review of applying deep learning on multi-modal sentiment analysis, particularly with regard to the integration of textual, visual and audio data to better interpret sentiments. Combining financial time series analysis with deep learning methods has seen interest as an ability to capture temporal dynamics. Chen et al. [23] surveyed deep learning models for standalone and hybrid models for financial time series prediction, and concluded that integrated approaches for sentiment-driven financial forecasting are more effective in financial forecasting. In addition, as for this utility of hybrid feature extraction techniques, such as Word2Vec-TFIDF, which combines Word2Vec with Term Frequency-Inverse Document Frequency (TFIDF), George and Murugesan [24] showed their utility in enhancing the analysis of sentiment in financial news headlines.

As such, Mohamed et al. [25] developed a hybrid lexicon and deep learning model called LexDeep to analyze unemployment-related Twitter discourse on COVID-19. Finally, this model serves to demonstrate that hybrid methods are adaptable to domain-specific applications such as financial crises. In sector-level sentiment analysis, Almalis et al. [26] extended use cases of deep learning to financial markets and showed how granular insights in the financial markets may be obtained. Furthermore, Ewald and Li [27] also studied how news sentiment contributes to forecasting commodity prices, thus filling the gap between financial forecasting and sentiment analysis.

3 Methodology

The method used in this study is centered around a mechanical framework designed to investigate and analyze the research question.

3.1 CNN

CNN is a type of neural network that is specifically designed to cater to the format of feed data, images in the form of photographs or any format that is arranged in a matrix [19]. CNN uses convolutional filters to obtain feature hierarchies from the input it receives. In the studies relying on the sentiment analysis of financial language, CNN has been applied to extract all semantic and syntactic features from the input texts [2, 20]. The convolutional layer develops its own filters to build up feature maps which are further passed on to the pooling layer to get the important features. Several convolutional-pooling layers allow the extraction of the features up to alternative abstraction levels. It has been observed that many features generated out of CNN contain a lot of useful discriminative information that would help in achieving a very accurate sentiment classification.

3.2 SVM

SVM is one of the most commonly used machine learning algorithms that is employed in different fields. It is a kind of supervised technique for learning by training that is applicable in cases of classification as well as

regression. SVM is the type of supervised machine learning model commonly used for the purpose of classification tasks [9]. It is a type of classification algorithm with the objective of looking for the so-called hyperplane to fully and representatively separate the diverse classes. Each hyperplane is placed in such a way that maximizes the distance between the support vectors – the points which are most distant from the separating boundary. SVM has known capacity to respond to intricate decision boundaries and multi-class conditions appropriately [10]. When it comes to sentiment analysis, SVM can be used for performing the multi-class classification. This classification process involves employing features extracted from CNN to categorize financial text into three sentiment polarities. It is either positive, negative or neutral [4]. Optimization of the SVM, together with the concept of large margin separation, leads to improved generalization performance and the ability to classify new unseen data samples.

3.3 RF

RF, also referred to as random decision forest, is a machine learning algorithm that is made up of many decision trees.

RF is another kind of classification model based on the family of ensemble models constructed with a number of decision trees [11]. In each decision tree, the features of the learning dataset are randomly selected for training the tree. Invoking the case of sentiment classification, the features that are extracted from CNN are used as features by multiple decision trees. They also do independent multi-class sentiment classification with each tree. The macro-level sentiment polarity is then determined through voting or averaging of the forecasts from all trees [13]. In fact, RF employs many uncorrelated models in order to strengthen the accuracy and robustness of the forecast, which is better than that of an independent decision tree. The application of ensemble methods allows for the idea of gaining stability when there is noise inside complex textual data.

3.4 Hybrid Models

In this concept, hybrid models are described as a type of computational model that consists of many techniques or methods that are used in solving a specific problem or task.

(a) CNN-RF hybrid model

CNN-RF is a synergistic model between CNN and RF. The hybrid model combines CNN and RF classifiers [6, 14]. Firstly, for the CNN, convolutions and pooling methods are used to obtain HFRs by entering other financial text data. RF classifier inputs are characteristics that incorporate semantic knowledge. RF is a set of decision trees that work independently, but with a goal of multi-class sentiment classification. For the rating-sentiment, the overall polarity of the sentiment is computed from the individual trees by voting or averaging [14]. The hybrid CNN-RF model integrates the feature learning technique of CNN and the ensemble technique of RF to enhance the performance of sentiment analysis.

(b) CNN-SVM hybrid model

SVM-CNN is one of the frequently used deep learning models, which includes CNN with SVM in its framework [6, 20]. In the CNN module, textual features are obtained by convolving the input text. The capturing characteristics of semantic relationship properties are then forwarded to an SVM classifier. In this case, SVM is applied to the task of multiple-class classification. It does this by using trained decision hyperplanes to identify the separation of sentiment into the positive, negative and neutral categories. This model combines the feature-extracting skills of CNN with the general skills of SVM for sentiment classification [20].

3.5 Model Parameter Settings

(a) CNN architecture

- Embedding layer: The vocabulary size (vocab_size) was set to 10,000 words, the embedding dimension was set to 128, and the input length (max_length) was fixed at 200.
- Convolutional layer: The number of filters was set to 64, the kernel size was set to 5, and ReLU was used as the activation function.
- Pooling layer: Global Max Pooling was employed as the pooling type.
- Dense layers: The first dense layer consists of 64 neurons with ReLU activation. The output layer is composed of three neurons, corresponding to the sentiment classes, with the use of a softmax activation function.

(b) Optimizer and loss function

- Optimizer: Adam was employed as the optimizer.
- Learning rate: The default learning rate from the Keras Adam optimizer was used.
- Loss function: Sparse Categorical Crossentropy was used as the loss function for multi-class classification.
- Metrics: Accuracy was used as the evaluation metric.

(c) Training settings

- Epochs: The model was trained for 100 epochs.
- Batch size: A batch size of 64 was used.

- Validation split: 20% of the dataset was allocated for validation.

3.6 Network Architecture

The architecture of the network was defined as follows: input → embedding layer → Conv1D → global max pooling → dense layer (ReLU) → dense output layer (Softmax).

CNN is used both as a classifier and a feature extractor. For feature extraction, all layers except the final dense layer are retained.

3.7 Optimization Strategies

(a) CNN optimization

Early stopping can be implemented to prevent overfitting. In addition, hyperparameter tuning can be considered, such as the learning rate, filter size, and kernel size, via grid search or random search.

(b) Feature-based models

- RF: The number of trees (n_estimators) was set to 100, and the random state was fixed at 42 for reproducibility.
- SVM: The kernel was set to linear, and the regularization parameter (C) was set to 1.0.

3.8 Data Preprocessing

(a) Text data preprocessing

- Tokenization: Text was converted into integer sequences using the Tokenizer from Keras.
- Padding: Padding was applied to ensure uniform sequence length, with a maximum length (max_length) of 200.
- Label encoding: String labels were converted into integers using LabelEncoder.

(b) Dataset splitting

- Train-test split: The dataset was split into 80% for training and 20% for testing.
- Random state: A random state of 42 was used.

3.9 Evaluation Metrics

- Accuracy: Accuracy was used to evaluate the overall performance.
- Classification report: The classification report included precision, recall, and F1-score for each class.
- Confusion matrix: The confusion matrix was used to visualize prediction errors.
- Plots: Plots of accuracy/loss over epochs and sentiment distributions were generated to provide insight into model behavior.

4 Results

In this study, two hybrid models, CNN + RF and CNN + SVM, were developed for financial sentiment analysis and compared for their performance. The analysis is based upon examining validation accuracy and validation loss with respect to varied training configurations in order to highlight the strengths and weaknesses of each model.

4.1 Performance of the CNN + RF Model

(a) Model configuration and results

Table 1 shows the configuration and performance results of the CNN + RF model.

Table 1. Configurations of the CNN + RF model

Configuration	Epochs	Batch Size	Validation Accuracy	Validation Loss
Configuration 1	150	64	0.6801	2.2488
Configuration 2	100	64	0.6869	2.0131
Configuration 3	50	64	0.6818	1.9538

(b) Performance summary

The highest validation accuracy (0.6869) was achieved by Configuration 2. The lowest validation loss (1.9538) occurred in Configuration 3.

4.2 Performance of the CNN + SVM Model

(a) Model configuration and results

Table 2 shows the configuration and performance results of the CNN + SVM model.

(b) Performance summary

The highest validation accuracy (0.6938) was achieved by Configuration 2. The lowest validation loss (1.7226) also occurred in Configuration 2, outperforming all CNN + RF configurations.

Table 2. Configurations of the CNN + SVM model

Configuration	Epochs	Batch Size	Validation Accuracy	Validation Loss
Configuration 1	150	128	0.6801	2.3228
Configuration 2	100	64	0.6938	1.7226
Configuration 3	50	64	0.6903	1.8402

4.3 Key Observations

(a) Comparison of models

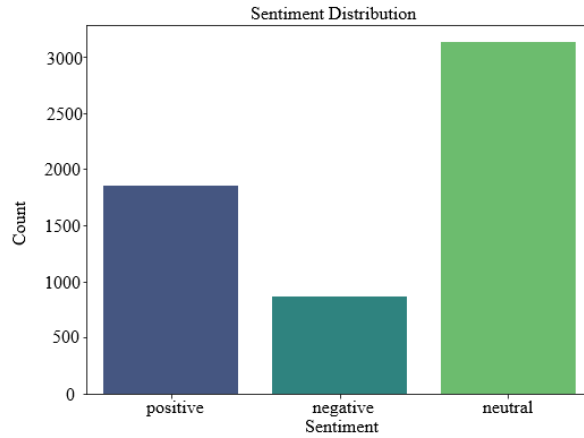
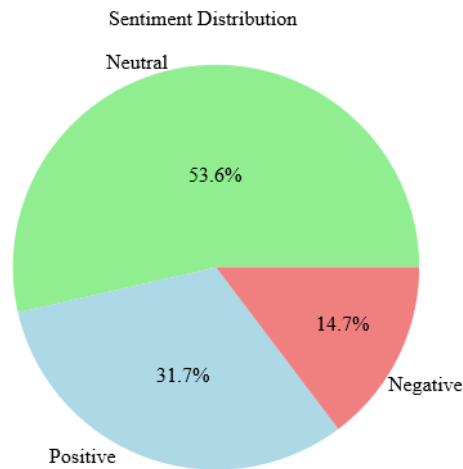
CNN + SVM consistently outperformed CNN + RF in both validation accuracy and validation loss. The highest validation accuracy of CNN + SVM (0.6938) exceeded that of CNN + RF (0.6869). The lowest validation loss of CNN + SVM (1.7226) was significantly better than CNN + RF (1.9538).

(b) Overfitting

Training accuracy for CNN peaked at 0.92 after 100 epochs, with a corresponding loss of 0.10. However, validation accuracy decreased from 0.69 to 0.67, and validation loss increased to 2.37, indicating overfitting.

4.4 Model Superiority

CNN + SVM (Configuration 2) outperformed other configurations, yielding the highest accuracy and the lowest loss. As seen in Table 1, the proposed CNN + RF method has a slightly higher validation loss compared to the results of the SVM method, meaning that SVM is more appropriate in this case.

**Figure 1.** Sentiment distribution (CNN+RF)**Figure 2.** Pie chart sentiment distribution (CNN+RF)

4.5 Insights on Hybrid Models

Combining the deep learning model (CNN) with conventional machine learning models (RF and SVM) has proved to be more efficient. The combination of CNN and SVM turned out to be relatively easy to implement and highly effective for the task of financial sentiment analysis.

The final validation accuracy achieved was 0.666, which is pretty close to the test accuracy of 0.666. The other method is alignment, which, in addition to this performance graph, further supports the presence of overfitting. The model performed so much better on the training data versus unseen data. The performance was evaluated against various metrics, including precision, recall, and F1-score. The overall performance was found satisfactory, but areas for improvement remain. The class 0 has the worst performance of the classes in the confusion matrix by frequently misclassifying. This can be provoked by imbalance between classes. CNN + RF model displays an obvious sign of overfitting on the whole since it has a lackluster metric difference from training metrics to validation as well as test metrics. The performance of the RF classifier is also not good enough to compensate for the overly fitted features of the CNN which would lead the model to be very hard to generalize.

The figures can help learn how the training and evaluation of the model. Figure 1 and Figure 2 show that the dataset has a large class imbalance, with positive sentiment being the minority class (30.8%) and neutral sentiment being the majority class (53.6%). Careful data preprocessing is required due to this imbalance since the model is likely to end up predicting an unbalanced dataset of ratios.

The confusion matrix of CNN + RF is shown in Figure 3 and has better performance in identifying neutral sentiment than negative sentiment. This discrepancy indicates that the model needs to be refined to give better minority class performance.

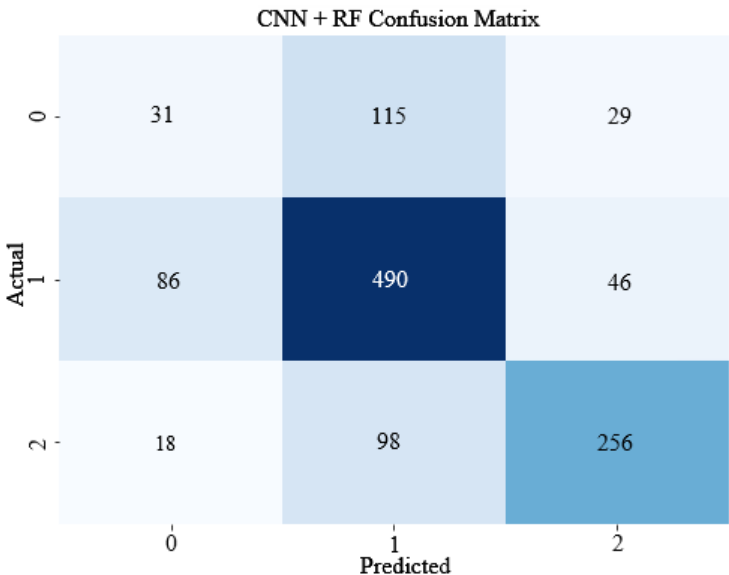


Figure 3. Confusion matrix (CNN+RF)

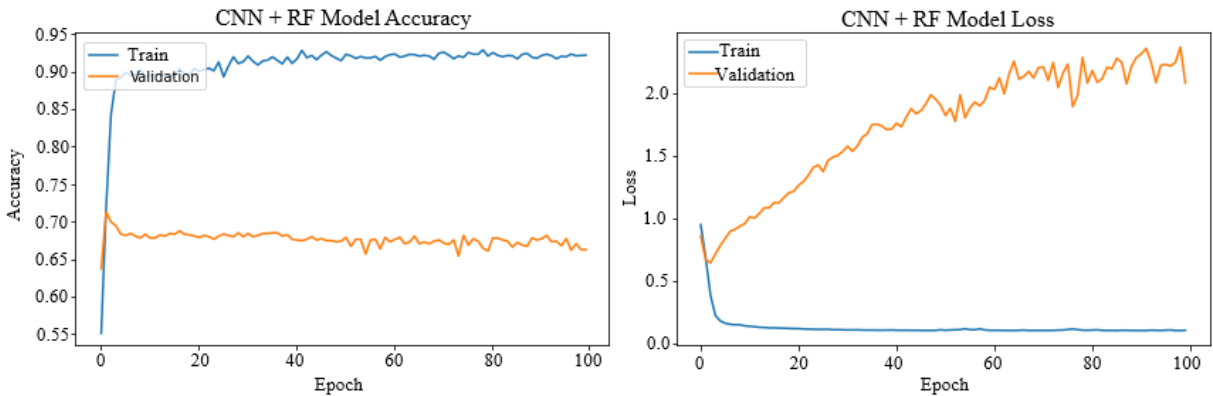


Figure 4. Model accuracy (CNN+RF)

The challenge of overfitting with the CNN + RF model is also demonstrated in Figure 4. Finally, training accuracy and loss were steadily improved, but validation metrics revealed a large discrepancy. Such overfitting points to the value of using strategies like regularization, dropout, or reducing model complexity to improve generalization.

In order to improve the performance of CNN, advanced regularization techniques, such as increased L2 regularization and dropout layers, were suggested. The result showed that these methods can solve the issue of overfitting, i.e., poor generalization on unseen data when the model becomes too tailored to the training data.

Other potential improvements are as follows:

a) Addressing class imbalance: To ensure a better class distribution for the classes inside the dataset to improve the model performance for all classes.

b) Hyperparameter optimization for RF: The performance of the RF model can be fine-tuned once by passing parameters like `n_estimators` (number of trees) and `max_depth`.

c) Use of CNN as a feature extractor: Using the CNN to obtain features and then feeding them to a linear model like SVM instead of RF for better results.

Finally, overfitting was proven to still be the major source that limits the performance in this research. The sentiment analysis was performed in the financial domain using a hybrid CNN-SVM model. The approach suggests a need for a systematic improvement in feature extraction, hyperparameter tuning, and class handling for better results.

For classification imbalance in the dataset used, Figure 5 and Figure 6 show the effect of class imbalance on the model classifying minority classes (negative sentiments).

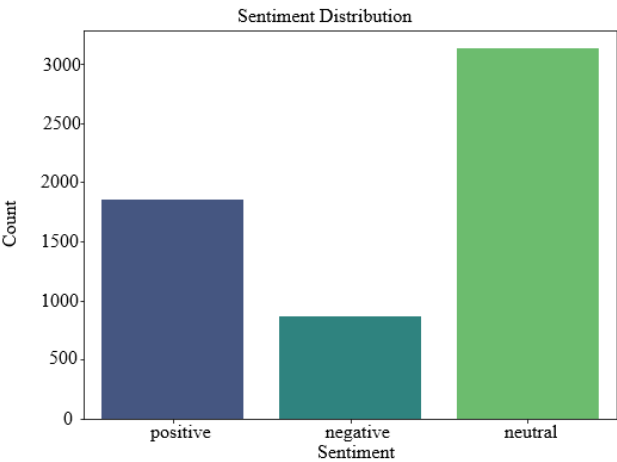


Figure 5. Sentiment distribution (CNN + SVM)

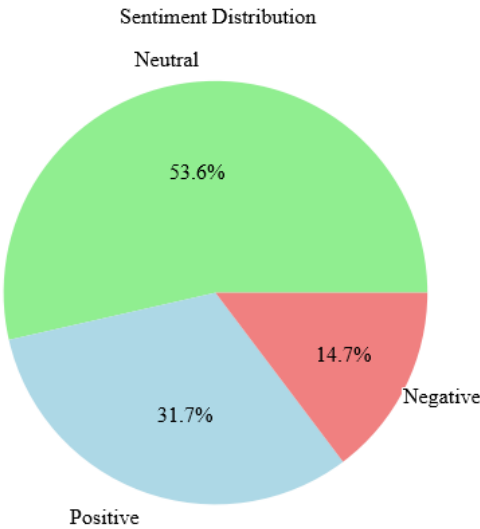


Figure 6. Pie chart sentiment distribution (CNN + SVM)

As shown in Figure 7, the CNN-SVM model does relatively better on neutral and positive classes. However, there is still room to improve for the negative sentiments.

As illustrated by Figure 8, overfitting is a major challenge, and the only few ways to correct it for the model to generalize properly are regularization, dropout, or introducing more data in the dataset.

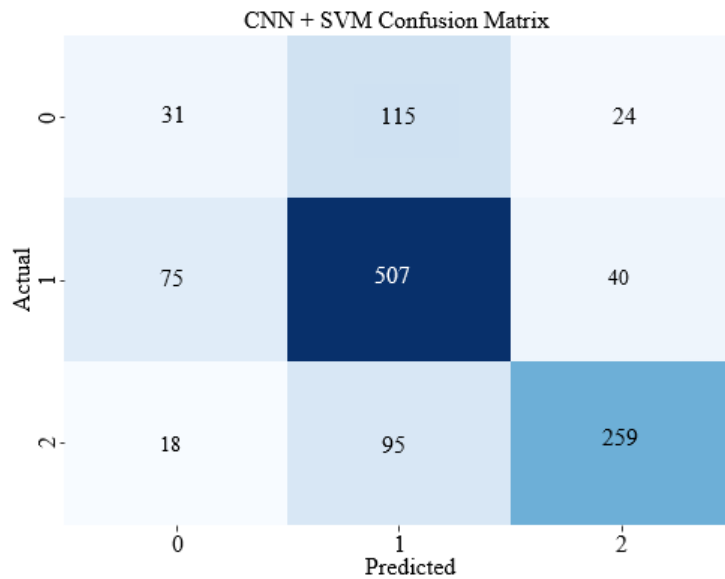


Figure 7. CNN+SVM confusion matrix

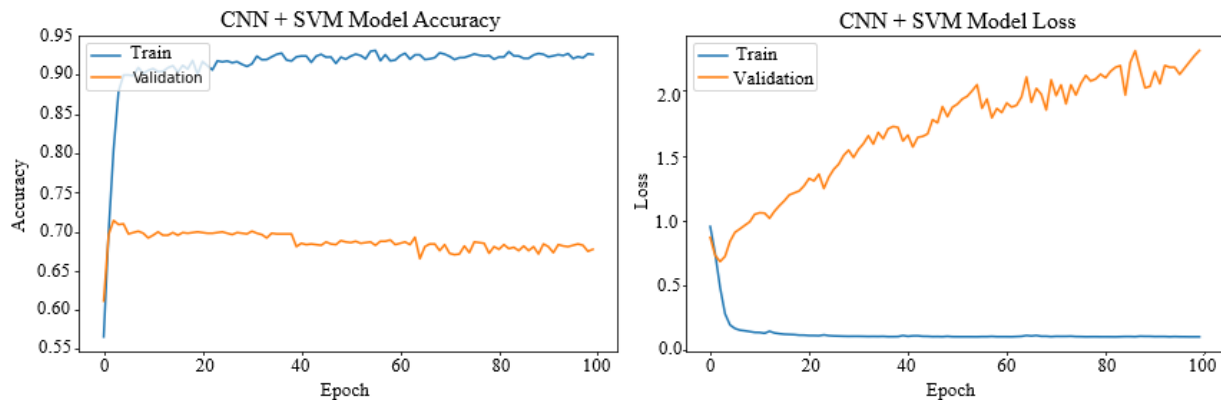


Figure 8. CNN+SVM model accuracy

According to Table 3, the key observations are as follows:

a) Validation accuracy

CNN + SVM always was better than CNN + RF in validation accuracy and had greater accuracy compared with CNN + RF (0.6938 vs. 0.6869).

b) Validation loss

CNN + SVM achieved lower validation loss (1.7226) compared to CNN + RF (1.9538), indicating better generalization.

c) Class-wise performance

The performance of both models was comparable with classification for the neutral class, but struggled in terms of negative sentiments. Classification of the minority classes was marginally better with CNN + SVM.

d) Overfitting

Both models were overfitting, but CNN + SVM showed slightly better validation performance (i.e., better overfitting resistance).

e) Robustness

In all, CNN + SVM turned out to be the overall more robust model because of higher validation accuracy and lower loss.

The CNN model achieved test accuracy of 0.686, which matches well to the validation accuracy and is lower than training accuracy. It reinforces that there is overfitting present in the model. Furthermore, further insight on the classification report was presented with metrics, i.e., precision, recall, and F1-score, for each class. The overall performance is adequate, but still a great deal of improvement is possible. It can be observed from the confusion matrix that the most problematic thing is with predicting class 0. It seems that many samples belonging to that class are misclassified as the other classes, which is probably due to the poor class imbalance characteristics of the dataset. To overcome these difficulties, several strategies were suggested, such as applying more robust regularization methods, coping better with class imbalance, as well as further tuning the hyperparameters, for better overall and generalization performance. Although improvements would not likely outperform, it would likely produce better predictions on minority classes and reduce the disparity of training and validation metrics observed currently.

Table 3. Comparison of both models

Aspect	CNN + RF	CNN + SVM
Dataset distribution	Both models faced the same class imbalance, with neutral sentiments dominating (53.6%).	The same dataset distribution challenge, with similar performance variations by sentiment class.
Best validation accuracy	0.6869 (Configuration 2 with 100 epochs and a batch size of 64).	0.6938 (Configuration 2 with 100 epochs and a batch size of 64).
Best validation loss	1.9538 (Configuration 3 with 50 epochs and a batch size of 64).	1.7226 (Configuration 2 with 100 epochs and a batch size of 64).
Confusion matrix insights	Better performance on the neutral class but struggled with minority classes (negative sentiment).	Slightly better performance on minority classes (negative sentiments) than CNN+RF.
Training accuracy	Reached about 92% during training.	Similar, reaching about 92% during training.
Overfitting behavior	Overfitting observed, validation accuracy plateaued, and validation loss increased over epochs.	Overfitting observed but slightly better validation performance compared to CNN+RF.
Performance trends	Stronger on neutral class; higher validation loss indicates less robust generalization.	Higher validation accuracy and lower validation loss indicate better generalization; better at minority class handling.

5 Conclusions

The results indicate that the proposed CNN-SVM and CNN-RF hybrid models are capable of having significantly better generalization capabilities than standalone approaches. The CNN-SVM model showed consistent improvements over the traditional methods in multi-class sentiment classification tasks by using CNN for robust feature extraction and SVM for handling complex decision boundaries. Likewise, the CNN-RF hybrid model was able to tackle high-dimensional and noisy datasets by utilizing the ensemble presented by RF, thereby improving in general the reliability of sentiment predictions. In this study, two key innovations were introduced, i.e., relying on the accuracy in extracting intricate semantic patterns of CNN coupled with the classification robustness of SVM and RF. These approaches can successfully deal with long-standing issues in financial sentiment analysis, including overfitting and class imbalance. The presented hybrid frameworks outperform existing models that often fail on these tasks with respect to both accuracy and interpretability.

These hybrid architectures were analyzed comparatively with previously published research, highlighting the gains achieved. In addition, even though standalone CNN, amongst other deep learning models, can provide very powerful feature extractors, it lacks the precision and stability of using traditional classifiers. This gap was filled by the proposed methods that achieve improved generalization and robustness and remediate these critical limitations of existing sentiment analysis techniques. The broader implications of this study are in using the developed features to advance sentiment analysis techniques in financial contexts and to understand the behavior of financial asset data. The hybrid models help investors, analysts and organizations make better decisions by allowing the sentiment classification to be more reliable. In future work, adversarial training might be added to further improve model robustness or possibly pre-trained domain-specific models. For instance, FinancialBERT could yield good performance in more niche financial datasets. The applicability of these directions can also extend hybrid approaches to dynamic financial environments and further refine such hybrid approaches.

Funding

This work is supported by the Scientific Research Project Fund of FIRAT ÜNİVERSİTESİ under the project number MF.24.46.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014. <https://doi.org/10.48550/arXiv.1408.5882>
- [2] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," *arXiv preprint arXiv:1510.03820*, 2015. <https://doi.org/10.48550/arXiv.1510.03820>
- [3] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara, "Deep learning for stock prediction using numerical and textual information," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, Okayama, Japan, 2016, pp. 1–6. <https://doi.org/10.1109/ICIS.2016.7550882>
- [4] Z. Hu, W. Liu, J. Bian, X. Liu, and T. Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, New York, NY, USA, 2018, pp. 261–269. <https://doi.org/10.1145/3159652.3159690>
- [5] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Using structured events to predict stock price movement: An empirical investigation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1415–1425. <https://doi.org/10.3115/v1/D14-1148>
- [6] X. Ding, Y. Zhang, T. Liu, and J. W. Duan, "Deep learning for event-driven stock prediction," in *Proceedings of the 24th International Conference on Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 2327–2333. <https://dl.acm.org/doi/10.5555/2832415.2832572>
- [7] J. F. Si, A. Mukherjee, B. Liu, Q. Li, H. Y. Li, and X. T. Deng, "Exploiting topic-based Twitter sentiment for stock prediction," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 2013, pp. 24–29. <http://doi.org/10.13140/2.1.3604.7043>
- [8] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9603–9611, 2015. <https://doi.org/10.1016/j.eswa.2015.07.052>
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995. <https://doi.org/10.1007/BF00994018>
- [10] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proceedings of 10th European Conference on Machine Learning*, Chemnitz, Germany, 1998, pp. 137–142. <https://doi.org/10.1007/BFb0026683>
- [11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001. <https://doi.org/10.1023/A:1010933404324>
- [12] H. Widiputra, A. Mailangkay, and E. Gautama, "Multivariate CNN-LSTM model for multiple parallel financial time-series prediction," *Complexity*, vol. 2021, no. 1, p. 9903518, 2021. <https://doi.org/10.1155/2021/9903518>
- [13] X. Liu, J. H. Guo, H. Wang, and F. Zhang, "Prediction of stock market index based on ISSA-BP neural network," *Expert Syst. Appl.*, vol. 204, p. 117604, 2022. <https://doi.org/10.1016/j.eswa.2022.117604>
- [14] D. M. Nelson, A. C. Pereira, and R. A. De Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, 2017, pp. 1419–1426. <https://doi.org/10.1109/IJCNN.2017.7966019>
- [15] Y. Yang, M. C. S. UY, and A. Huang, "FinancialBERT: A pre-trained language model for financial text mining," *arXiv preprint arXiv:2006.08097*, 2020. <https://doi.org/10.48550/arXiv.2006.08097>
- [16] M. Wen, P. Li, L. F. Zhang, and Y. Chen, "Stock market trend prediction using high-order information of time series," *IEEE Access*, vol. 7, pp. 28 299–28 308, 2019. <https://doi.org/10.1109/ACCESS.2019.2901842>
- [17] X. D. Li, H. R. Xie, L. Chen, J. P. Wang, and X. T. Deng, "News impact on stock price return via sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 14–23, 2014. <https://doi.org/10.1016/j.knosys.2014.04.022>
- [18] F. L. Feng, H. M. Chen, X. N. He, J. Ding, M. S. Sun, and T. S. Chua, "Enhancing stock movement prediction with adversarial training," *arXiv preprint arXiv:1810.09936*, 2019. <https://doi.org/10.48550/arXiv.1810.09936>
- [19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. <https://doi.org/10.1109/5.726791>

- [20] Y. T. Peng and H. Jiang, "Leverage financial news to predict stock price movements using word embeddings and deep neural networks," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, 2016, pp. 374–378. <https://doi.org/10.18653/v1/N16-1041>
- [21] M. Wankhade, C. Kulkarni, and A. C. S. Rao, "A survey on aspect base sentiment analysis methods and challenges," *Appl. Soft Comput.*, vol. 167, no. Part A, p. 112249, 2024. <https://doi.org/10.1016/j.asoc.2024.112249>
- [22] A. Pandey and D. K. Vishwakarma, "Progress, achievements, and challenges in multimodal sentiment analysis using deep learning: A survey," *Appl. Soft Comput.*, vol. 152, p. 111206, 2024. <https://doi.org/10.1016/j.asoc.2023.111206>
- [23] W. S. Chen, W. Hussain, F. Cauteruccio, and X. Zhang, "Deep learning for financial time series prediction: A state-of-the-art review of standalone and hybrid models," *Comput. Model. Eng. Sci.*, vol. 139, no. 1, pp. 187–224, 2024. <https://doi.org/10.32604/cmcs.2023.031388>
- [24] M. George and R. Murugesan, "Improving sentiment analysis of financial news headlines using hybrid Word2Vec-TFIDF feature extraction technique," *Procedia Comput. Sci.*, vol. 244, pp. 1–8, 2024. <https://doi.org/10.1016/j.procs.2024.10.172>
- [25] A. Mohamed, Z. M. Zain, H. Shaiba, N. Alturki, G. Aldehim, S. Sakri, S. F. M. Yatin, and J. M. Zain, "LexDeep: Hybrid lexicon and deep learning sentiment analysis using Twitter for unemployment-related discussions during COVID-19," *Comput. Mater. Contin.*, vol. 75, no. 1, pp. 1577–1601, 2023. <https://doi.org/10.32604/cmc.2023.034746>
- [26] I. Almalis, E. Kouloumpris, and I. Vlahavas, "Sector-level sentiment analysis with deep learning," *Knowl.-Based Syst.*, vol. 258, p. 109954, 2022. <https://doi.org/10.1016/j.knosys.2022.109954>
- [27] C. O. Ewald and Y. Y. Li, "The role of news sentiment in salmon price prediction using deep learning," *J. Commodity Markets*, vol. 36, p. 100438, 2024. <https://doi.org/10.1016/j.jcomm.2024.100438>