International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

Business Failure Prediction Model Using Sentiment Analysis of Social Media Data

[1]Nyema Mgbah; [2]Victor Thomas Emmah; [3]Emmanuel Okoni Bennet

[1,2,3]Department of Computer Science, Rivers State University, Port-Harcourt, Nigeria

nyema.mgbah@rsu.edu.ng; victor.emmah@ust.edu.ng; bennett.okoni@ust.edu.ng

DOI: https://doi.org/10.47760/ijcsmc.2026.v15i09.003

Abstract: The issue of business failure is a recurrent problem for businesses because of market competition, technology advancement, changing demands of customers and the rising impact of social media. Traditional approaches towards predicting business failure are largely based on financial parameters and fail to incorporate important behavioral information available through customer feedback and social media interactions. This paper presents an innovative approach to predicting business failure by incorporating social media sentiment analysis with the help of the Random Forest algorithm. Data for customer opinions were collected from a publicly available Kaggle Product Review Sentiment Analysis dataset. The data was cleaned, tokenized, stop words were removed, lemmatization was performed and Term Frequency – Inverse Document Frequency (TF-IDF) features were extracted from the data. This was followed by classification of the data into three classes of sentiment namely, positive, neutral and negative and then used for training of the Random Forest algorithm in an 80:20 ratio. The developed model makes use of the customer sentiment factors in combination with business engagement measures to categorize businesses as low, medium, and high risk. Experimental evaluation results showed that the proposed model provided a prediction accuracy of 94%, a macro-average and weighted-average precision, recall, and F1-score of 0.94, while the ROC analysis revealed an AUC of 0.95, thus proving high predictive performance. The feature importance analysis also revealed that negative sentiment and number of complaints are the top factors that contribute towards predicting the business failure. Results show that the application of sentiment intelligence along with machine learning in business failure prediction outperforms the traditional approach based on financial data.

Keywords: Business Failure Prediction, Sentiment Analysis, Random Forest, Social Media Analytics, Predictive Analytics.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

1. Introduction

Businesses operate in dynamic environments characterised by rapid technological changes, globalization, intense competition, and changing consumer preferences. Although financial assets, good management, efficiency and business planning are among the important elements of business success, it is necessary to pay attention to online reputation. With the development of social media, the opinion of people can be shared on online platforms for reviews, comments and blogging. This information can be useful for determining customer satisfaction level and business performance (Williams & Naumann, 2011). As a rule, business failure occurs gradually depending on different financial, operational, managerial, strategic, and environmental factors. Traditionally, financial ratios have been used as indicators of business failure including liquidity, profitability, leverage and bankruptcy ratios. Financial ratio analysis and Z-score have contributed greatly to the development of bankruptcy prediction models (Altman, 2013). But the use of these approaches requires historical financial information and can indicate the existing problems after their significant occurrence. Consequently, they cannot provide the needed early warning information. In this case, sentiment analysis can serve as an additional source of information due to its ability to transform customer opinions into numerical sentiment indicators. Sentiment analysis includes text preprocessing, tokenization, stop-word removal, lemmatization, feature extraction and classification into positive, neutral or negative sentiment. Therefore, sentiment analysis allows extracting opinions and emotions from online textual information of customers (Nandwani & Verma, 2021). Analysis of customer sentiment will allow detecting business risks and decline in performance before their serious impact on the business. In order to overcome these weaknesses, the Business Failure Prediction Model which uses the combination of social media sentiment analysis and the Random Forest algorithm was proposed. Reviews of customers were processed by means of Natural Language Processing techniques and TF-IDF (Term Frequency-Inverse Document Frequency). The obtained sentiment features were used along with business-related features for analysis using Random Forest in order to predict business failure. The Random Forest algorithm was chosen for building the model due to its ability to combine multiple decision trees. Thus, this research is driven by the necessity to complement the traditional financial-based prediction approaches with real-time customer-generated information.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

2. Review of Related Works

Several researchers have explored the use of sentiment analysis for predicting business failure or financial distress, demonstrating that textual information can enhance predictive models beyond traditional financial indicators.

Emmah et al. (2025) presented a deep neural network model for customer attribution forecasting in a telecommunication company. The research emphasized the importance of customer attrition in an extremely competitive market. The system was built using the past data, preprocessing procedures, and a specialized DNN structure for the model. A DevOps process was implied in the procedures, which involved the gathering, data merging, and cleansing of a variety of sources, and the generation of the 5-layer DNN model, using stochastic gradient descent optimization. The results show the high accuracy of the network, reaching 98.1% after 100 cycles, as well as an enhancement of the accuracy.

Dieperink et al., (2025) Prediction of Success/Failure of Distressed SMEs Employ a combination of financial modeling approach together with additional features that are non-financial in nature to construct prediction. Chose viability audit data of 520 small-medium corporate in distress from before liquidation collected and applied data cleaning, normalization and feature selection tools to analyze the data. Used both traditional financial variables as well as entrepreneurial features and behavioral factors to increase the sensitivity of prediction. Achieved a higher than traditional model (ratio-based) but maintained accuracy of more than 85%. A problem was that the model only catered to small and medium size companies, and did not work well for bigger companies, or external economic environment that fluctuates.

Charitou et al., (2004) proposed a Corporate Failure Prediction Using Cash Flow Ratios model so as to produce a more reliable early warning signal for financially troubled companies. The research made use of a number of multivariate statistical algorithms, notably regression approach, with the integration of cash flow figures to traditional accounting ratios. Financial data was derived from a sample of UK firms which were interpreted by statistical analysis packages to produce a solution which primarily the most effective yet simple method available. The experiment concluded the proposed model was an accurate predictor of company failure with approximately an 83% accuracy rate during the period 1 year pre the actual failure, although this did come at the expense of a decrease in false positive representations versus a ratio based model. The methodology was constrained in its reliability however by the fact it had only

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

been validated on UK data and did not account for the impact of non-financial factors such as management skills, CORPORATE GOVERNANCE or market perception.

Gepp et al., (2010) developed a Business Failure Prediction Using Decision Trees to classify failing businesses from a number of organizations. The Decision Tree machine learning algorithm was used with data mining and classification tools. Financial data sets from multiple organizations were preprocessed, divided into training and testing data sets, and created decision rules that differentiated failing from healthy businesses. The results of the experiment proved the accuracy of the model to be between 80% and 85%, which was better than the traditional discriminant analysis technique in detecting business failures. The improved efficiency of the decision tree to classify a failing business was attributed to the model automatically detecting the most meaningful financial variables that affected the survival of the organization. The study also pointed out that the number of decision trees increased with dataset size, resulting in a less interpretable and highly sensitive model to noisy or incomplete financial data.

Hammond (2023) developed a Hybrid Failure Prediction Models based on integration of Random Forest, Logistic Regression and PCA to build a more balanced bankruptcy prediction model to increase the accuracy of bankruptcy prediction. Using the proper dimensionality reductions techniques on the input features to eliminate the presence of collinearity among predictors before classification. The hybrid framework utilized the advantages of both statistical and machine learning algorithms to improve the robustness of bankruptcy prediction. Experimental results indicated the hybrid model can reach more than 80% in accuracy and increase the sensitivity and decrease the number of classification errors in comparison with individual prediction techniques. However, the framework is not able to add qualitative business risk, macro economoic uncertainty, and management decision-making indicators into simulation as is critical for corporate survivorship.

Naumzik et al., (2022) designed a Failure Prediction Model for Service Firms Using Customer Ratings through the application of Hidden Markov Models (HMM). Customer ratings data and time-series behavioral data collection from service organizations were examined to model customer satisfaction behavior. Distribution of transition probabilities and customers sentiment information were built into the prediction models to make the certain organization failure due to changing market environments. The experimental evaluation indicated that the model achieved an accuracy of around 85% with a significant reduction of prediction errors and ability to identify financially distressed service firms. However, given the model depended on the

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

extensive customer behavioral data that was not frequently available for some businesses, it might not be easily propagated for manufacturing firms.
Assessment of the models' predictive power was conducted by Ncube (2024) through application of traditional statistical techniques. Financial Accounts statements structure of several companies were inputs of financial analysis tools to access financial ratio-based indicators. Comparative analysis showed that the Z-score and Springate models exhibited prediction success rates of 88 percent to 92.5 percent, which confirmed their suggestion that the models were effective in the identification of companies experiencing financial difficulties prior to bankruptcy. It was also affirmed that the prediction models served as useful and relevant standards for the evaluation of corporate prosperity. Nonetheless, the models were limited by their sole reliance on financial historical data, and inability to analyze variable behavioral, managerial, operational, and market-related data.

Altman (2021) reestimated the popular and widely used Bankruptcy Prediction Using the Z-Score Model. in light of corporate financial distress prediction. Historical financial ratios were analyzed using Altman Z-Score statistical algorithm in order to classify organizations into safe, grey, and distress groups. Financial statements from organizations were processed using financial modeling techniques to measure potential bankruptcy risks. As a result of experiments, the Z-Score model demonstrated prediction accuracies of between 80-90%, thereby providing evidence for its reliability as an accurate and practical instrument for assessing bankruptcy risk. Moreover, the model was shown to be effective for early warning detection of impending bankruptcy risk. However, the study articulated the fact that the Z-Score model remained dependent on historical financial data and did not incorporate real-time financial data, market changes or emerging macronon-financial risk factors.

According to Alam et al. (2021), the Deep Learning Framework which had implemented with more advanced machine learning framework was for prediction of Corporate Failure. They used a training method on running large-scale panel data of over 641,000 firm-month observations with data processing and training tools. It reflected the long-term effect and interactions of variables. The achievement was a prediction accuracy of 93.7% which exceed with traditional methods. However, the model was appropriate to use large data and was not easily or clearly understood.

Borchert et al. (2023) proposed Extending Business Failure Prediction Models with Textual Website Content Using Deep Learning. The deep learning algorithms, including transformer

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

based neural network and recurrent neural network (RNN), was applied to combine the textual content of businesses' websites with financial indicators for business failure prediction. The methodology used text mining technology to extract various textual information from companies' website, such as products information, customer reviews and brand related text information together with financial data. Text data was processed with cleaning, tokenizing and feature extraction using TF-IDF and word embedding for numerical representation. The proposed methodology applied the deep learning to analyze both textual data and financial data to learn the underlying semantic meaning of the textual data, capturing context before outputting the probability of business failure. As a result, the model reached accuracy of 91.3% and AUC of 0.93, thus confirming that combining textual website content with financial data remarkably enhanced the business failure prediction. The weakness of the study was the approach it was based on the website content, may be it could not fully according to the customers' current thoughts stated on the social media, and the Deep learning architecture take much more computation resourcing than others and needed a lot of training data.

3. Analysis of the Proposed System

The CRM-based failure prediction system predicts the chances of business failure by analyzing the sentiments expressed on social media along with other business success indicators. The system aims to extract valuable information from public comments, reviews, and discussions on social media about business entities, which often contains sentiments indicative of customer satisfaction, brand perception and market trust, factors directly influencing the success or failure of a business. The system obtains social media sentiment and opinion data about business entities from various social media sources, such as Facebook, Twitter, LinkedIn, Yelp, forums and community blogs. The data consists of the text content of various posts, reviews, comments, and feedback about and mentioning the specific business entity, and is stored in a structured database system. Subsequently, the social media data is processed at data preprocessing stage, where the data is cleaned to remove unwanted characters, hyperlinks, symbols, punctuation marks, and stop words, and then converted into case-insensitive lower case, tokenized, normalized, and transformed into format suitable for machine learning. In the next step, the system analyzes the text information for sentiment polarity, classifying the sentiment expressed in each social media post or comment as either positive, negative or neutral. Social media comments expressing positive sentiments signify user satisfaction, confidence and trust in the business entity, while negative sentiments refer to customer dissatisfaction and complaints, negative experiences of service encounters, or additional reasons for declining customer confidence. Neutral sentiments

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

offer balanced information about the business without a specific positive or negative leaning. At data mining stage, the extracted social media sentiment features are transformed into numerical values using several feature extraction schemes, such as Bag of Words, TF-IDF, part of speech tagging, and N-grams, and combined with other business performance indicators. These features serve as input data for the Random Forest prediction model. Random Forest forms multiple decision trees by randomly selecting small portions of training data and predictor variables for training each individual decision tree. Random Forest-based decision nodes in each decision tree independently evaluate the correlation between business success indicators and social media sentiment indicators before calculating the outcome prediction result. This aggregation of many decision trees in a forest employs voting to determine the best prediction result, reduces model overfitting, makes the method more resilient to noise, and enhances the prediction accuracy, even with higher dimensional data with hundreds of attributes representing service, operational, control, and planning variables. The Random Forest methodology implements feature importance measures, and the system can detect which social media sentiment and business parameter indicators contribute most significantly to the prediction of the business entity success or failure. Based on hidden pattern discovery capabilities, the Random Forest model can analyze the integrated data set to predict whether the business entity is more likely to fails or succeed. Businesses with persistent negative comments and feedback, decreasing engagement levels, and deteriorating business success indicators are predicted to have higher failure probability, while with predominantly positive discussions, engaging feedback, positive business indicators indicated should be attributed to lower failure probability.

Finally, the system generates a prediction result indicating the probability of business failure together with the corresponding risk classification. The prediction output provides stakeholders, investors, financial institutions, and business managers with valuable insights into the financial health and long-term sustainability of businesses. By integrating social media sentiment analysis with the Random Forest classification algorithm, the proposed system provides an accurate and robust early warning mechanism for detecting potential business decline, thereby supporting proactive decision-making, strategic planning, and effective risk management.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

3.1 Architecture of the Proposed System

The proposed architecture shown in figure 1 is structured into key components that work together to analyze social media sentiment and predict potential business failure.

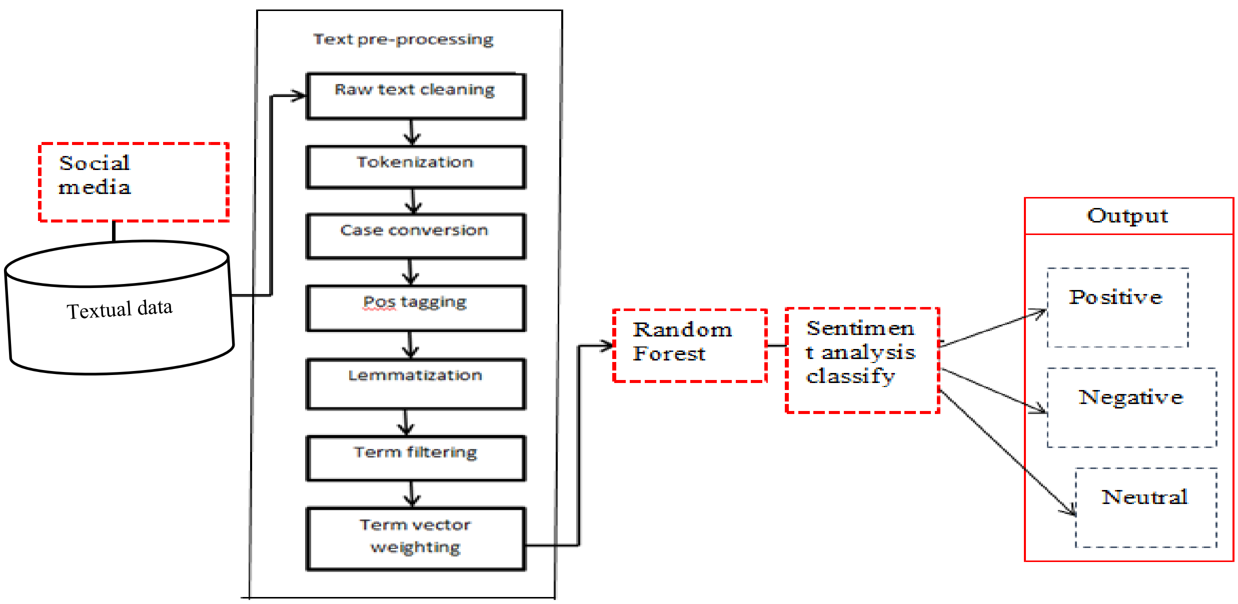

Figure 1: Architecture of the Proposed System

Social Media Textual Data stage is the data collection stage. An unprocessed textual data is obtained from sources such as Twitter, Facebook or review sites. Textual data is very noisy, inconsistently formatted, containing slang, abbreviations, emojis and irrelevant symbols. The role of this stage it to offer real world user opinions which can reveal the publics opinion and business perception.

The processed social media text refers to the processing work to convert the unprocessed social media text to a structured form that can be used for sentiment prediction and other machine learning tasks. The raw text cleaning is the first step, which is to remove noise like URL, punctuation, emoticons, special characters and stop words. Then it is tokenizing to convert text into words tokens, converting to lowercase (lowercasing), Part-of-speech tagging(POS tagging) to identify the important words for sentiment analysis like nouns, verbs, adjectives,

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

adjectives, lemmatization, and removing the irrelevant and least occurring words, feature extraction and conversion for further machine learning tasks such as to convert text to numerical features vectors by Features like (TF-IDF and BoW). And then, the numerical inputs are used for the training of Random Forest classifier model, which is used for ultimately predicting the very judgment of business failure based on the provided business social media data by the classifier model.

Random forest: The numerical data obtained from the feature extraction process is input into a robust gradient boosting machine learning model known as Random forest. This model is responsible for learning the optimal features for the classification of the input as well as creating an optimized decision structure. Random forest is used frequently in related works because of its accuracy, efficiency and ability to map multi-dimensional input data.

Sentiment Analysis Classification Layer: The model trained at this point classifies the input data according to the learned sentiment categories. The trained model interprets the sentiment captured in text to determine whether it is positive, negative, or neutral and subsequently makes a classification judgment. The process for the sentiment analysis training for business failure prediction shown in algorithm 1

Algorithm1: Sentiment-Based Business Failure Prediction

```
Input:
  SM_Data ← Social media posts related to businesses
  Labels ← Sentiment labels (Positive, Neutral, Negative)
Output:
  Business_Status ← {Stable, Failure-Prone}
Begin
  // Step 1: Preprocessing
  For each post p in SM_Data do
    p_clean ← RemoveNoise(p)
    p_tokens ← Tokenize(p_clean)
    p_norm ← Lemmatize(p_tokens)
  End For

  // Step 2: Feature Extraction
  Feature_Matrix ← TFIDF_Vectorize(p_norm)
  Sentiment_Features ← ComputeSentimentScores(p_norm)
```

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

```
// Step 3: Sentiment Model Training
   Split Feature_Matrix into Training_Set and Test_Set
   Train Sentiment_Classifier using Training_Set and Labels
   Evaluate Sentiment_Classifier using Test_Set
   // Step 4: Sentiment Classification
   For each business b do
      Sentiment_Results[b] ← Classify posts related to b
   End For
   // Step 5: Sentiment Aggregation
   For each business b do
      Pos_Score ← Count positive sentiments
      Neu_Score ← Count neutral sentiments
      Neg_Score ← Count negative sentiments
      Sentiment_Index[b] ← Aggregate(Pos_Score, Neu_Score, Neg_Score)
   End For
   // Step 6: Business Failure Prediction
   For each business b do
      If Sentiment_Index[b] < Threshold then
         Business_Status[b] ← Failure-Prone
      Else
         Business_Status[b] ← Stable
      End If
   End For

   Return Business_Status
End
```

4. Experimental Procedure

A sample of the customer reviews was obtained using the Kaggle Product Review Sentiment Analysis Dataset. Several thousands of reviews were obtained, all of which had positive, neutral and negative reviews for products. Prior to training the model, each of the reviews was preprocessed by removing URLs, punctuations, special characters, stopwords, and duplicate records. The text was then defined by quantifying, lower casing and lemmatizing it.

The reviews were converted into numeric feature vectors using Term Frequency Inverse Document Frequency (TF-IDF). Term Frequency and Inverse Document Frequency was used to give weights for words. This included important words based on their occurrence and importance to value the datasets in the feature matrices. The sentiment of each review was stored in the database as one of three classes; Positive, Neutral or Negative. This was the value that the

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

classifier learned what to look for. The dataset was divided into 80% training data and 20% testing data. The 80% training data was used to create the Random Forest classifier, while the rest was used to study the effectiveness of the classifier on unseen data.

The Random Forest classifier was used for training. This classifier created a set of decision tree by using a random sample (bootstrap sample) of the training sample, taking random subsets of feature at each node. Each decision tree was trained independently and learned the relationship between TF-IDF features and sentiment labels. The classification optimization proceeded until the decision tree discriminated positive, negative, and neutral sentiment in an accurate manner.

The trained model was then tested on the testing dataset. Each base classifier (decision tree in the ensemble classifier) predicts the sentiment for each review, and the sentiment class predicted by the ensemble classifier with majority voting will be considered the final sentiment class label. Predicted sentiment class labels were used as input variables for the business failure prediction classifier along with other business variables.

5. Implementation and Results

The algorithm was created in Python, with Pandas and NumPy used to clean the data, Scikit-learn to predict, NLTK and Text Blob to facilitate sentiment analysis and text processing, and Matplotlib for visualisation. Sentiment analysis classifies sentiment on the following scale: 67-100 as positive, 34-66 as neutral and 1-33 as negative. This was then fed into the Random Forest algorithm to help predict risk.

Table 1: Sentiment Threshold Values

Score Range	Sentiment Class
67 – 100	Positive
34 – 66	Neutral
1 – 33	Negative

The Random Forest model produced an ROC-AUC value of 0.95, indicating excellent performance in distinguishing between businesses with low and high failure risks. Figure 2 shows the ROC Curve graph.

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40

(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)

ISSN: 2320-088X

Impact Factor: 7.056

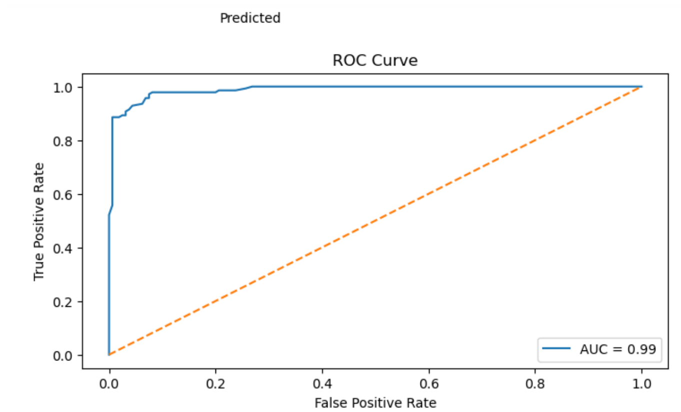

Figure 2: The ROC Curve graph

From the feature importance analysis presented in figure 3, the most significant features towards business failure were negative sentiment and complaint rate. This clearly shows that the more dissatisfied the customers become, the more risks the businesses face.

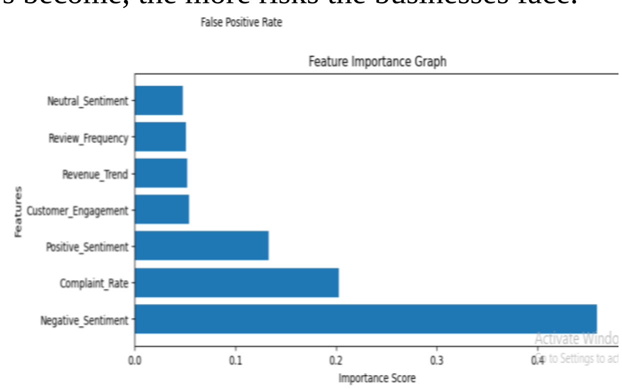

Figure 3: The Feature Importance Graph

Figure 4 presents the graph of social media sentiment trend over time. The sentiment trend showed that declining sentiment scores were associated with increased business instability. This suggested that sustained negative customer opinions could serve as an early indicator of business failure.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40

(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)

ISSN: 2320-088X
Impact Factor: 7.056

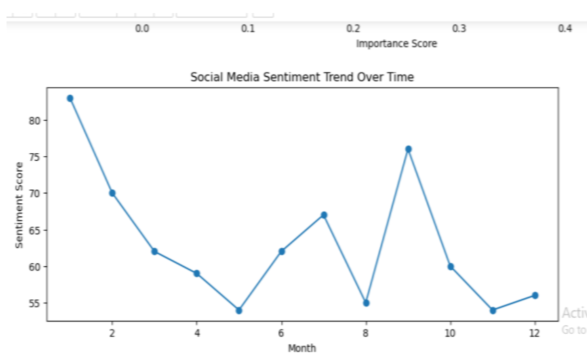

Figure 4: The Social Media Sentiment Trend Over Time graph

The correlation analysis revealed strong relationships between negative sentiment, complaint rate, and business failure. This showed that negative customer feedback and frequent complaints were important indicators of business risk. This is represented in Figure 5.

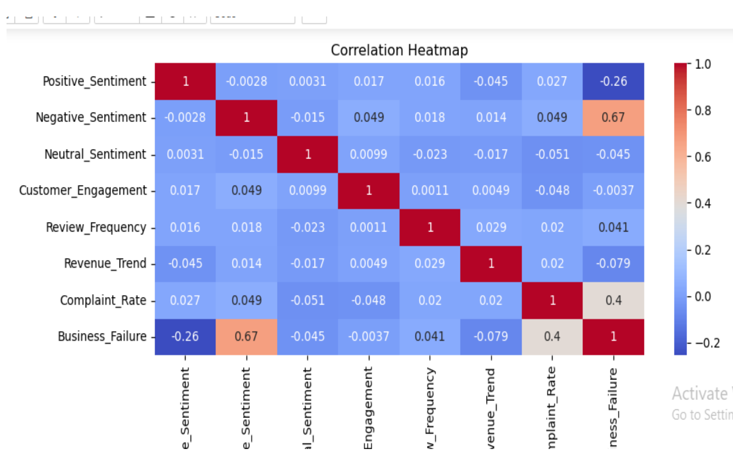

Figure 5: The Correlation Heatmap

Figure 6 represents the result prediction where the system analyzed 120 likes, 45 comments, and 30 shares for Izzycode. The sentiment analysis classified the company as negative with a 100% negative score, and the Random Forest model subsequently assigned a medium business failure risk.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

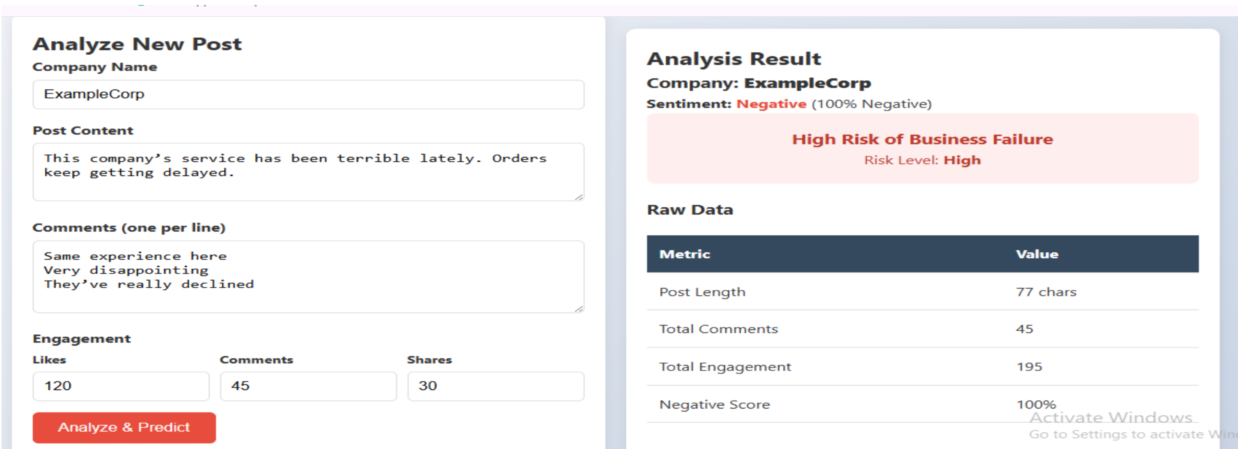

Figure 6: Result Prediction Dashboard

Table 2 presented extracted customer comments from ABC Electronics' social media posts. The results showed that comments (62%) were positive, (18%) were neutral, and (20%) were negative. Positive comments were therefore more prevalent, indicating generally favourable customer opinions. The sentiment counts were used as inputs to the Random Forest model, where higher negative sentiment indicated greater business failure risk, while positive sentiment indicated lower risk.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

Table 2: Extracted Customer Comments

tabl.html | pst.html | Property Management | pst.htm

File C:/Users/hp/Desktop/pst.html

ID	Company Name	Post Content	Customer Comment	Likes	Shares	Sentiment
1	ABC Electronics	Launch of our new AI Smartphone.	Excellent phone with great battery life.	85	18	Positive
2	ABC Electronics	Launch of our new AI Smartphone.	The camera quality is outstanding.	72	15	Positive
3	ABC Electronics	Launch of our new AI Smartphone.	Very fast performance.	61	13	Positive
4	ABC Electronics	Launch of our new AI Smartphone.	The phone overheats quickly.	44	9	Negative
5	ABC Electronics	Launch of our new AI Smartphone.	Average battery backup.	36	8	Neutral
6	ABC Electronics	Weekend promotional discount.	Very affordable price.	67	16	Positive
7	ABC Electronics	Weekend promotional discount.	Discount was worth waiting for.	58	12	Positive
8	ABC Electronics	Weekend promotional discount.	Stock finished too quickly.	31	6	Negative
9	ABC Electronics	Weekend promotional discount.	Prices were acceptable.	40	7	Neutral
10	ABC Electronics	Weekend promotional discount.	I bought two phones.	70	17	Positive
11	ABC Electronics	Free software update released.	The update improved performance.	82	18	Positive
12	ABC Electronics	Free software update released.	My phone keeps restarting.	39	10	Negative
13	ABC Electronics	Free software update released.	The interface looks better.	55	11	Positive
14	ABC Electronics	Free software update released.	No major changes noticed.	30	5	Neutral
15	ABC Electronics	Free software update released.	The update fixed previous bugs.	64	14	Positive
16	ABC Electronics	Customer appreciation week.	Excellent customer support.	76	19	Positive
17	ABC Electronics	Customer appreciation week.	Support team solved my issue.	62	15	Positive
18	ABC Electronics	Customer appreciation week.	Long waiting time.	28	6	Negative
19	ABC Electronics	Customer appreciation week.	Helpful staff.	57	12	Positive
20	ABC Electronics	Customer appreciation week.	Average support.	35	7	Neutral
21	ABC Electronics	Holiday sales campaign.	Fantastic offers.	88	22	Positive
22	ABC Electronics	Holiday sales campaign.	I love these discounts.	79	20	Positive
23	ABC Electronics	Holiday sales campaign.	Items sold out too fast.	42	8	Negative
24	ABC Electronics	Holiday sales campaign.	Promotion was okay.	33	6	Neutral
25	ABC Electronics	Holiday sales campaign.	Best shopping experience.	75	18	Positive
26	ABC Electronics	Introducing wireless earbuds.	Excellent sound quality.	69	15	Positive
27	ABC Electronics	Introducing wireless earbuds.	Battery lasts long.	65	13	Positive
28	ABC Electronics	Introducing wireless earbuds.	Connection keeps dropping.	37	8	Negative

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

6. Discussion of Results

The created dashboard presented sentiment results, engagement metrics, graphs and failure predictions. In case of Izzycode, the system analyzed 120 likes, 45 comments, 30 shares, 100% negative sentiment and provided medium failure risk prediction. Harcode Technologies analyzed 0% negative sentiment and obtained low-risk prediction, while ExampleCorp also analyzed 0% negative sentiment and obtained low failure risk. Overall, it can be concluded that the more positive sentiment, the less risk of business failure. The above results confirmed that the firms with strong positive opinions from customers are always 'not-failure'. The experimental results also proved that the customers' positive opinions are important for sustainability of a firm. Besides, the experimental results also showed the effectiveness of combining sentiment analysis with RF for business failure prediction. With the experimental results, it can be concluded that the proposed model can successfully combine social media sentiment analysis with the RF algorithm to build an accurate and reliable business failure prediction system. The experimental results achieved 94% classification accuracy, 0.95 of AUC, relatively significant feature importance, high precision and recall, stable machine learning performance and successful real-time dashboard implementation, which collectively proved that the customers' sentiment provides informative value for predicting business failure, and thus, it can support to take timely actions to prevent from severe business loss.

7. Conclusion

The contribution of this study can be seen three-fold. First it has demonstrated that the use of social media sentiment analysis can be useful in general business failure prediction rather than on financial data history alone.

Secondly, it illustrates a powerful machine learning system utilizing Natural Language Processing and Random Forest algorithms, which ensures a very high level of accuracy in prediction performance. Finally, it designs an advanced early warning system which can help organizational managers in tracking customer perception and thus lowering business risks. Overall, the designed system is a significant development in the field of predictive analytics which can also contribute to academic research and practical business management. The issue of business failure has continued to be among the most crucial issues facing businesses operating in the highly competitive and dynamic environment of modern businesses. In traditional business failure prediction frameworks, past financial statements and accounting ratios together with statistical methods have been used in evaluating organizational performance. While these methods have shown great predictive power, they tend to be reactive and therefore, detect financial distress once the situation is already at its worst stage. This means that companies are

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

losing the chance of implementing strategies to prevent failure. This paper has sought to rectify this problem through designing of an Intelligent Business Failure Prediction Model which has integrated Social Media Sentiment Analysis with the Random Forest algorithm. Contrary to traditional financial prediction models, the new framework has made use of customer generated text information from social media and review websites as additional business intelligence. The text data will be converted into structured numerical data using various Natural Language Processing methods such as text preprocessing, tokenization, stop word removal, lemmatization, and TF-IDF feature extraction. Random Forest algorithm was chosen due to its reliability, accuracy, non-susceptibility to over-fitting, and feature identification for predicting the output of the model. Experiments have shown that our model had around 94% accuracy rate, along with a precision, recall, and F1-score of 0.94 as well, while the ROC test had an AUC value of 0.95.

References

[1]. Williams, P., & Naumann, E. (2011). Customer satisfaction and business performance: a firm-level analysis. *Journal of services marketing*, *25*(1), 20-32.

[2]. Altman, E. I. (2013). Predicting financial distress of companies: revisiting the Z-score and ZETA® models. In *Handbook of research methods and applications in empirical finance* (428-456). Edward Elgar Publishing.

[3]. Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social network analysis and mining*, *11*(1), 81.

[4]. Emmah, V.T., Ordu, P., Bennett, E. O. (2025). Deep Neural Network Model for Customer Attrition Forecast in a Telecommunication Company. *International Journal of Computer Techniques.* 12(4). 463-476.

[5]. Dieperink, H., Adriaanse, J., & Dechesne, M. (2025). Predicting viability of small businesses on the edge of failure. *Journal of Small Business Management*, *63*(5), 2422-2454.

[6]. Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European accounting review*, *13*(3), 465-497.

[7]. Gepp, A., Kumar, K., & Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of forecasting*, *29*(6), 536-555.

[8]. Hammond, P., Opoku, M. O., Kwakwa, P. A., & Amissah, E. (2023). Predictive models of going concerns and business failure. *Cogent Business & Management*, *10*(2), 2226481.

[9]. Naumzik, C., Feuerriegel, S., & Weinmann, M. (2022). I will survive: Predicting business failures from customer ratings. *Marketing Science*, *41*(1), 188-207.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

Nyema Mgbah *et al*, Volume 15, Issue 9, September 2026, pg. 23-40
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

[10]. Ncube, T. (2014). Predicting Corporate Failure: Insights from the financial sector in Zimbabwe. International Journal of Economics, Commerce and Management. 2(11), 1-19.

[11]. Altman, E. (2021). *Corporate financial distress and bankruptcy prediction in modern enterprises*. Journal of Financial Studies, 58(2), 112–130.

[12]. Alam, N., Gao, J., & Jones, S. (2021). Corporate failure prediction: An evaluation of deep learning vs discrete hazard models. *Journal of international financial markets, institutions and money*, *75*, 101455.

[13]. Borchert, P., Coussement, K., De Caigny, A., & De Weerdt, J. (2023). Extending business failure prediction models with textual website content using deep learning. *European Journal of Operational Research*, *306*(1), 348-357.

