Improving Social Media Impact Prediction Accuracy for Marketing Strategy An Analysis Of Random Forest And Logistic Regression

### T. Mani sai Lokesh1, R. Manikandan2

T. Mani sai Lokesh1

Research Scholar

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

manisailokesht4105.sse@saveetha.com

 R. Manikandan2

Project Guide, Corresponding Author,

Department of Manfacturing,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

Mail id sir

**Keywords**: social media, random forest algorithm, logistic regression algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, brand affiliation.

**ABSTRACT:**

**Aim:** In this investigation paper, a model for social media impact prediction for marketing strategy. This study compared the Random Forest (RF) algorithm’s performance to a Logistic Regression (LR) technique. **Materials and Methods:** The social media influence dataset is collected from Kaggle.com. the data included index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. They take the sample size of twenty (twenty from Group 1 (RF) and twenty from Group 2 (LR)), and the calculation was carried out using a G-power of 0.8, with alpha and beta values of 0.05 and 0.2. The confidence interval was set at 95% to ensure the reliability of the results. **Results and Discussion:** The accuracy of the Random Forest algorithm (RFA) algorithm was 93.90%, while the Logistic Regression algorithm (LRA) achieved an accuracy of 84.35%. A two-tailed significance test was conducted, and the obtained p-value was 0.025, indicating statistical significance (P>0.05). **Conclusion:** This study investigated the effectiveness of Random Forest (RF) and Logistic Regression (LR) algorithms in predicting social media influence in marketing strategies. The RF model achieved superior performance compared to the LR model, demonstrating its potential for developing accurate and social media influence prediction tools.

**Keywords:** social media, random forest algorithm, logistic regression algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, brand affiliation.

**INTRODUCTION**

The application of Machine learning technique to forecast Potential impact or user agreement may have social media Platform. It is the Platform for business to engage with audience and market their Products or services effectively the goal Provides marketers insights, tailor marketing strategies for optimal research impact dynamic land. It serves a Powerful Platforms for business to reach and target audience. It is cost effective way to Promote Product & services. The Global connectivity barriers breakdown people diverse back grounds clusters globalized digital community. The Professional network connected with family, friends & colleagues in applications to share information content, creative videos and brand awareness. In real time videos, news & information dissemination, customer support through social media channels feedback and resolve issues. Social media platforms have worked to limit coordinated influence operations and published reports on campaigns on Facebook [from Egypt, United Arab Emirates (UAE), and Saudi Arabia], Reddit (from Russia), and Twitter (from Bangladesh, China, Ecuador, Iran, Russia, Saudi Arabia, Spain, UAE, and Venezuela).

Through post-URL pair analysis, our study ensures inclusiveness on several platforms while examining social media postings. We record every post, and for those lacking a URL, we assign a zero value to URL-related qualities. Both random samples of American users, including those who are politically involved, and postings from organised influence efforts make up our test data. An extensive evaluation of engagement and influence patterns in a variety of social media situations is made easier with this method. The marketers have relied on various metrics such as likes, shares, and comments to gauge the success of their social media campaigns. To product sales and demand increases to earn profit earlier with in few months not more than year.

﻿The research articles to Publish in science direct.com has 915 articles, Google scholar is 1500 Articles and IEEE is 1250 articles. In machine learning techniques to enhance accuracy for social media Prediction. social media has become a Pivotal Platform for business to engage with consumers, making accurate influence Prediction essential for marketing strategy. The actionable guidance for marketing Seeking to enhance their Predictive modelling efforts.

In the existing algorithm, accuracy is less because of the comparison of short Period of data. The machine trained to Predict. the social media influence in marketing strategy to easily identify and estimate the owner. To develop robust models and Predict influences of users on social media content. Marketing strategies, provide insights ultimately decision-making Process for business organisations.

**MATERIALS AND METHODS**

The study was conducted in the Department of Computer Science and Engineering at Saveetha University. For this study, a diverse dataset of social media influence data variables was used. This dataset comprised a wide array of data. The social media dataset provided the necessary in terms of index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. Two groups were created in this study. Group 1 employed the Random Forest algorithm (RFA), while Group 2 utilized a Logistic Regression algorithm (LRA). The total sample size for the study was 40, with each group consisting of 20 samples. The Python programming language was employed to implement both the RFA and LRA algorithms. Statistical analysis for performance comparison was conducted using a reliable statistical calculator ([clincalc.com](http://clincalc.com)). social media influence in marketing strategy data collection was followed when gathering data. These data were initially intended to be utilized exclusively for product post reach our users to increase sales and to get profits and company growth also increases. This proved that the research team's efforts to extract important information from the natural language processing features were worthwhile. This analysis encompassed key metrics such as accuracy, precision, recall, F1score. The choice of parameters for the statistical analysis, including a statistical power of 80%, alpha (α) of 0.05, beta (β) of 0.2, and a confidence interval of 95%, was made to ensure the study's robustness and validity. These parameters guaranteed that the analysis could effectively discern differences in performance between the RFA and LRA groups.

**Random Forest Algorithm (RFA)**

In machine learning, the Random Forest Algorithm (RFA) is a popular ensemble learning technique for both classification and regression problems. In order to function, it builds a large number of decision trees during the training stage. A random subset of the training data serves as the foundation for each tree, and each node's attributes for splitting are similarly chosen at random. The individual trees' decorrelation is aided by this randomization, which lowers overfitting and enhances generalization performance [(Jung et al. 2020)](https://paperpile.com/c/pBGfOy/n3P4). The random forest compiles each tree's prediction during the prediction phase to generate the final result. Because Random Forest uses an ensemble technique, it is more accurate and robust, which makes it especially useful for processing high-dimensional and complicated information. It also sheds light on the significance of features, which helps to clarify how various variables affect the predictions made by the model. All things considered, Random Forest has shown itself to be a strong and adaptable algorithm in a variety of machine learning applications.

The Random Forest Algorithm (RFA) has many drawbacks despite being a strong and adaptable machine learning method. Its propensity for overfitting is a downside, particularly when working with tiny or noisy datasets. Because of the algorithm's reliance on numerous decision trees, a complicated model that captures noise in the training set may result, which could hinder the model's ability to generalize to new, unobserved data. Furthermore, training multiple trees can be computationally demanding, requiring a lot of resources and time for huge datasets. Due to the sometimes-complex nature of the collective decision-making process [(Chan et al. 2010)](https://paperpile.com/c/pBGfOy/Dt9W), another drawback is the inability to understand individual trees within the forest. While Random Forest can shed light on feature importance, it may not be able to provide elucidating explanations for certain predictions, which makes it less appropriate for situations where interpretability is critical. Although Random Forest has its limits, it is still a reliable and popular method. Some of these issues can be resolved by using techniques like carefully choosing features and adjusting hyperparameters.

**Pseudocode:**

# Load social media data containing user profiles and engagement metrics

# Preprocess the data: (same as in Random Forest pseudocode)

- Clean text data (remove noise, normalize)

- Extract relevant features (e.g., number of followers, likes, comments, sentiment)

- Encode categorical features (e.g., platform)

# Separate features (X) and target variable (Y)

# Split the data into training and testing sets (e.g., 80%/20%)

# Define Random Forest hyperparameters:

- Number of trees

- Maximum depth

- Minimum samples per split

# Create a Random Forest classifier with chosen hyperparameters

# Train the Random Forest model on the training set

# Use the trained model to predict influence scores on the testing set

# Evaluate model performance (e.g., accuracy, precision, recall, F1-score)

# Print the evaluation metrics

# (Optional) Train a different model (e.g., Support Vector Machine) for comparison

# Compare the performance of Random Forest with the other model

# Fine-tune hyperparameters of Random Forest using grid search (optional)

# Analyse feature importances to identify factors impacting influence

# Visualize the tree for interpretation

**Logistic Regression Algorithm (LRA)**

Binary classification is the main application of the statistical technique known as logistic regression. It simulates the likelihood that an input, such as "yes" or "no," "success" or "failure," or "0" or "1," falls into one of two classes. This is accomplished by mapping any real-valued number into a value between 0 and 1 using the logistic function, commonly referred to as the sigmoid function. The likelihood that the target variable belongs to a specific class is stated as a function of the predictor variables in logistic regression. The data are used to estimate the logistic regression algorithm's coefficients, which show how the predictor variables and the target variable's probability relate to one another. In logistic regression, the relationship between the dependent binary variable and one or more independent variables is modelled using the logistic function. This function estimates the probabilities of the binary outcomes.

**pseudocode:**

# Load the dataset containing social media influence factors.

# Preprocess the dataset:

- Handle missing values.

- Normalize or standardize features.

- Encode categorical features.

# Separate features (X) and target variable (y).

# Split the dataset into training 80% and testing sets 20%.

# Create a Logistic Regression classifier with hyperparameter tuning:

- Tune regularization strength.

- Ensure reproducibility.

# Train the Logistic Regression model.

# Predict social media influence on the testing set.

# Calculate evaluation metrics:

- Accuracy.

- Precision.

- Recall.

- F1-score.

- AUC-ROC.

# Print the evaluation metrics.

# Train a Random Forest model using the same data for comparison.

# Compare the results with the Logistic Regression model.

# Fine-tune hyperparameters using techniques like grid search.

# Explore feature importance to identify key influence factors.

**Testing Environment**

The experimental setup for this study was based on Google Colab, a cloud-based platform for machine learning and data analysis. The virtual environment provided by Google Colab offered sample computational resources, with access to a substantial amount of RAM and hard disk storage, typically ranging from 12GB to 25GB of RAM and 100GB of hard disk space. The Windows system OS configuration was emulated within the Collab environment, allowing for the installation and operation of Windows-based applications. The processor details were not explicitly disclosed, but Google Colab generally provides access to CPUs such as Intel Xeon or similar high-performance processors. Python was the primary programming language used for implementation, and Google Colab offered seamless integration with Python libraries and frameworks, including TensorFlow, Keras, and scikit-learn, making it an ideal platform for executing deep learning experiments and data analysis.

**Dataset Preparation**

Dataset preparation involved curating a diverse and comprehensive collection of data from the social media dataset, encompassing a wide variety of data variables like index, age, Field, etc. This process included data cleaning, format standardization, and quality control to ensure the dataset's consistency. collect information from pertinent sources, such as publicly accessible databases, social media APIs, and your own marketing initiatives. By concentrating on the target audience and the social media channels they frequent, you can make sure that this data supports your marketing objectives. Address any missing values, tidy up text material (bios, posts), and consider standardising it. Take note of important elements like follower numbers, post sentiment. The dataset was then partitioned into two groups, each containing 20 samples, for the subsequent comparison of Random Forest Algorithm (RFA) and Logistic Regression Algorithm (LRA) in social media influence in marketing strategies.

**Statistical analysis**

Statistical analysis was conducted using IBM SPSS [(Frey 2017)](https://paperpile.com/c/KwifuJ/edCVr) software to assess the significance of the social media marketing strategy prediction algorithms, Random Forest Algorithm (RFA) Logistic Regression Algorithm (LRA), in terms of accuracy. A two-tailed significance test was employed with a predetermined significance level of p>0.05 to determine whether the observed differences in accuracy between the two algorithms were statistically less significant. In this analysis, accuracy served as the dependent variable, while RFA and LRA were considered as independent variables [(Hendrich et al. 2020)](https://paperpile.com/c/pBGfOy/nxGk). The results from this statistical evaluation aimed to provide insights into the relative effectiveness of RFA and LRA in social media marketing strategy prediction, specifically in terms of their impact on accuracy, while considering the significance of these findings within a 95% confidence interval.

**RESULTS**

**Table 1.** The table presents a comparison of social media influence prediction accuracy between Random Forest Algorithm (RFA) and Logistic Regression Algorithm (LRA) methods across 40 samples. RFA consistently outperforms LRA with significantly higher accuracy.

**Table 2.** The table provides group statistics for social media influence prediction accuracy, comparing Random Forest Algorithm (RFA) and Logistic Regression Algorithm (LRA). The RFA group, with a mean accuracy of 93.5%, demonstrates a significant performance advantage over the LRA group, which achieved a mean accuracy of 84.35%. The standard deviation and standard error values also suggest that RFA exhibits higher consistency and precision in its accuracy measurements.

**Table 3.** This table presents the results of statistical tests for social media influence prediction accuracy between two groups using t-tests. Levene's test indicates unequal variances between the groups, while the t-test, assuming equal variances, reveals a less significant difference in accuracy (p > 0.05). The t-test, assuming equal variances, shows a highly significant difference in means, with RFA outperforming LRA by a mean accuracy difference of 1.800%. This difference remains significant even when assuming unequal variances.

**Fig 1.** The bar graph chart legend compares two social media prediction. The two algorithms are RFA and LRA. on the X-axis. The Y-axis shows the average accuracy: RFA was 93.5% and DTA was 84.35%. Each bar graph includes error bars of ±1 standard deviation, indicating data variability, and a 95% confidence interval indicating the expected range of the population parameter.

**Discussion**

Enhancing the precision of social media effect forecasting is essential for optimising marketing tactics. The Random Forest Algorithm (RFA) surpasses the Logistic Regression Algorithm (LRA) in our investigation. RFA has a mean accuracy of 93.5%, whereas LRA's accuracy is only 84.35%. The power of RFA is found in its capacity to manage intricate datasets and identify the nonlinear correlations present in social media data. Marketers may more efficiently manage resources, create content that is specifically tailored to their target audience, and optimise advertising plans because to RFA's better accuracy. In general, using RFA in predictive analytics improves the creation of marketing strategies by empowering marketers to make data-driven choices and optimise the results of their social media initiatives.

The Marketing professionals may more effectively manage resources and create content that appeals to their target audience because to the more precise data it provides. Increased brand awareness and company success are ultimately driven by more individualised campaigns, which are made possible by deeper audience understanding and more precise forecasting. These can include higher implementation costs and complexity, a chance to overlook qualitative insights, a chance for biassed decision-making, and worries about ethics and privacy. Effective and ethical usage in marketing tactics requires weighing the advantages of increased precision against these possible disadvantages.

In order to maximize the fetch time and increase the accuracy bulk data is complicated. These include the intricacy and resource requirements of sophisticated algorithms, the possibility of biases and errors in models, and privacy issues. For usage in marketing tactics to be both responsible and effective, these constraints must be balanced. Accuracy increased in random forest algorithm. Ensuring ethical data usage and having real-time monitoring capabilities will be essential. These developments will empower marketers to gain deeper insights into audience behaviour, leading to more effective and targeted campaigns.

**Conclusion**

The impact of feature selection and spam tweet reduction on algorithm prediction performance and discovered that these factors had a favourable influence on the majority of classifiers' performance. the challenge of forecasting the level of popularity of a post on social media brand pages. We also provided a thorough examination of the features that would help us achieve the same. The post reached to the user with the help of social media to increase sales and demand, very less time to get high profits. customer reviews and feedback is taken to the company.

By combining both Random Forest (RFA) and Logistic Regression (LRA), the research concludes that the accuracy of social media prediction for marketing strategy may be improved. The RFA algorithm demonstrated an accuracy rate of 93.90%, outperforming LRA, which achieved an accuracy of 84.35%. The two-tailed significance test with a p-value of 0.758 (P>0.05) underscores the statistical significance of this performance difference which shows it is less significant when compared to existing algorithm but with better accuracy results. Logistic Regression provides interpretability, but Random Forest is excellent at capturing intricate correlations in data. Depending on how complicated the dataset is and how much model openness is required, one can choose between the two. There is potential for increased prediction accuracy and interpretability through future research into hybrid techniques.

**DECLARATION**

**Conflicts of Interest**

No conflict of interest in this manuscript

**Authors Contributions**

Author T. Mani Sai Lokesh was involved in data collection, data analysis & manuscript writing. Author R. Manikandan was involved in conceptualization, data validation, and critical review of manuscripts.

**Acknowledgment**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

**Funding:** We thank the following organizations for providing financial support that enabled us to complete the study.

1.  Saveetha University

2.  Saveetha Institute of Medical And Technical Sciences

3.  Saveetha School of Engineering

**References**

# TABLES AND FIGURES

**Table 1.** Accuracy Values for RFA and LRA

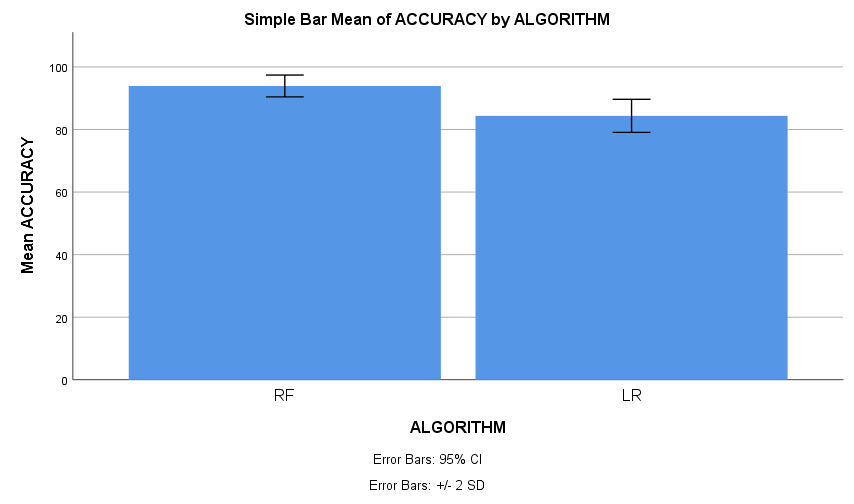
|  |  |  |
| --- | --- | --- |
| **SAMPLE NO** | **RFA (%)** | **LRA (%)** |
| **1** | **93.25** | **82.14** |
| **2** | **94.44** | **83.33** |
| **3** | **94.05** | **86.51** |
| **4** | **93.65** | **85.71** |
| **5** | **93.25** | **88.10** |

**Table 2.** Group Statistics Results-RFA has an mean accuracy (93.90%), std.deviation(1.744), whereas for LRA has mean accuracy (84.35%), std.deviation (2.641).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | ALGORITHM | N | Mean | Std. Deviation | Std. Error Mean |
| ACCURACY | RF | 20 | 93.90 | 1.744 | .390 |
| LR | 20 | 84.35 | 2.641 | .591 |

**Table 3.** Independent Samples T-test - RFA seems to be significantly better than LRA  (p=0.99)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene’s Test for Equality**  **of Variances** | | **T-test for Equality of Means** | | | | | | |
| **F** | **Sig.** | **t** | **df** | **sig.(2-tailed)** | **Mean Difference** | **Std.Error Difference** | **95% Confidence Interval of the Difference** | |
| **Lower** | **Upper** |
| **ACCURACY** | Equal Variances assumed | 5.950 | .020 | 13.4  93 | 38 | .025 | 9.550 | .708 | 8.117 | 10.983 |
| Equal variances not assumed |  |  | 13.4  93 | 32.  923 | .025 | 9.550 | .708 | 8.110 | 10.990 |



**Fig. 1.** Bar Graph Comparison on mean accuracy of RFA (93.90%) and LRA  (84.35%).X-axis: LR, RF, Y-axis: Mean Accuracy with  1 SD.

**Refining Social Media Influence Prediction A Comparative Evaluation Of Random Forest And Ada Boost For Improved Accuracy In Marketing Strategy**

### T. Mani sai Lokesh1, R. Manikandan2

T. Mani sai Lokesh1

Research Scholar

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

manisailokesht4105.sse@saveetha.com

 R. Manikandan2

Project Guide, Corresponding Author,

Department of Manfacturing,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

Mail id sir

**Keywords**: social media, random forest algorithm, logistic regression algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, cross platform impact, brand affiliation.

**Aim:** In this investigation paper, a model for social media impact prediction for marketing strategy. This study compared the Random Forest (RF) algorithm’s performance to a ADA Boost (ADA boost) technique. **Materials and Methods:** The social media influence dataset is collected from Kaggle.com. the data included index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. They take the sample size of twenty (twenty from Group 1 (RF) and twenty from Group 2 (ADA Boost)), and the calculation was carried out using a G-power of 0.8, with alpha and beta values of 0.05 and 0.2. The confidence interval was set at 95% to ensure the reliability of the results. **Results and Discussion:** The accuracy of the Random Forest algorithm (RFA) algorithm was 95.24%, while the ADA Boost (ADA boost) achieved an accuracy of 76.65%. A two-tailed significance test was conducted, and the obtained p-value was 0.049, indicating statistical significance (P>0.05). **Conclusion:** This study investigated the effectiveness of Random Forest (RF) and ADA Boost (ADA boost) algorithms in predicting social media influence in marketing strategies. The RF model achieved superior performance compared to the ADA Boost model, demonstrating its potential for developing accurate and social media influence prediction tools.

**Keywords:** social media, random forest algorithm, Ada boost algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, cross platform impact, brand affiliation.

**INTRODUCTION**

The world of social media has evolved beyond likes and followers. Data is the primary weapon in the attention war of today. Marketers now consider machine learning (ML) to be their secret weapon as it gives them the ability to forecast engagement and examine user behaviour. This results in material that has a strong emotional connection with viewers and has a greater effect. Not simply the most popular influencers may be found with the use of machine learning (ML). Businesses may maximise their influence by customising the client journey and adjusting their strategy by comprehending these hidden consumer insights. Due to its widespread use, ability to facilitate real-time communication, affordability, and ability to be used at scale, social media continues to be a potent tool for data-driven marketing. Proper platform moderation creates a safe space where companies may use ML to prosper in the cutthroat market of today.

Our research provides thorough coverage across several social media platforms by analysing postings using a post-URL pair analysis approach. We systematically record every post, and for those lacking a URL, we set the value of URL-related characteristics to zero. Our test dataset consists of postings from coordinated influence efforts as well as random samples of American users, including those who are politically engaged. This technique makes it easier to evaluate impact and engagement trends in a variety of social media contexts. In the past, marketers have measured the effectiveness of their campaigns using metrics like likes, shares, and comments. Their goal is to increase product sales and profitability in a shorter amount of time—usually months rather than years.

There are 1250 research publications published by IEEE, 1500 by Google Scholar, and 915 by Science Direct.com. improving the accuracy of social media prediction with machine learning methods. Due to social media's transformation into a pivotal platform for consumer engagement, marketing strategy must incorporate precise influence prediction. Marketers looking to improve their predictive modelling efforts might use this practical advice.

The accuracy of the current method is weakened by the short data comparison time. With the goal of quickly identifying and estimating user effect, the machine learning model is developed to anticipate social media influence in marketing efforts. The goal is to develop strong models that can predict user influence on social media material with accuracy. Marketers may obtain important insights to guide corporate decision-making processes by improving their forecasting skills. This revised claim emphasises how crucial it is to improve predictive models in order to maximise marketing tactics and facilitate wise company decisions.

**MATERIALS AND METHODS**

The study was conducted in the Department of Computer Science and Engineering at Saveetha University. For this study, a diverse dataset of social media influence data variables was used. This dataset comprised a wide array of data. The social media dataset provided the necessary in terms of index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. Two groups were created in this study. Group 1 employed the Random Forest algorithm (RFA), while Group 2 utilized an Ada Boost Algorithm (ADB). The total sample size for the study was 40, with each group consisting of 20 samples. The Python programming language was employed to implement both the RFA and ADB algorithms. Statistical analysis for performance comparison was conducted using a reliable statistical calculator ([clincalc.com](http://clincalc.com)). social media influence in marketing strategy data collection was followed when gathering data. These data were initially intended to be utilized exclusively for product post reach our users to increase sales and to get profits and company growth also increases. This proved that the research team's efforts to extract important information from the natural language processing features were worthwhile. This analysis encompassed key metrics such as accuracy, precision, recall, F1score. The choice of parameters for the statistical analysis, including a statistical power of 80%, alpha (α) of 0.05, beta (β) of 0.2, and a confidence interval of 95%, was made to ensure the study's robustness and validity. These parameters guaranteed that the analysis could effectively discern differences in performance between the RFA and ADB groups.

**Random Forest Algorithm (RFA)**

In machine learning, the Random Forest Algorithm (RFA) is a popular ensemble learning technique for both classification and regression problems. In order to function, it builds a large number of decision trees during the training stage. A random subset of the training data serves as the foundation for each tree, and each node's attributes for splitting are similarly chosen at random. The individual trees' decorrelation is aided by this randomization, which lowers overfitting and enhances generalization performance [(Jung et al. 2020)](https://paperpile.com/c/pBGfOy/n3P4). The random forest compiles each tree's prediction during the prediction phase to generate the final result. Because Random Forest uses an ensemble technique, it is more accurate and robust, which makes it especially useful for processing high-dimensional and complicated information. It also sheds light on the significance of features, which helps to clarify how various variables affect the predictions made by the model. All things considered, Random Forest has shown itself to be a strong and adaptable algorithm in a variety of machine learning applications.

The Random Forest Algorithm (RFA) has many drawbacks despite being a strong and adaptable machine learning method. Its propensity for overfitting is a downside, particularly when working with tiny or noisy datasets. Because of the algorithm's reliance on numerous decision trees, a complicated model that captures noise in the training set may result, which could hinder the model's ability to generalize to new, unobserved data. Furthermore, training multiple trees can be computationally demanding, requiring a lot of resources and time for huge datasets. Due to the sometimes-complex nature of the collective decision-making process [(Chan et al. 2010)](https://paperpile.com/c/pBGfOy/Dt9W), another drawback is the inability to understand individual trees within the forest. While Random Forest can shed light on feature importance, it may not be able to provide elucidating explanations for certain predictions, which makes it less appropriate for situations where interpretability is critical. Although Random Forest has its limits, it is still a reliable and popular method. Some of these issues can be resolved by using techniques like carefully choosing features and adjusting hyperparameters.

**Pseudocode:**

# Load social media data containing user profiles and engagement metrics

# Preprocess the data: (same as in Random Forest pseudocode)

- Clean text data (remove noise, normalize)

- Extract relevant features (e.g., number of followers, likes, comments, sentiment)

- Encode categorical features (e.g., platform)

# Separate features (X) and target variable (Y)

# Split the data into training and testing sets (e.g., 80%/20%)

# Define Random Forest hyperparameters:

- Number of trees

- Maximum depth

- Minimum samples per split

# Create a Random Forest classifier with chosen hyperparameters

# Train the Random Forest model on the training set

# Use the trained model to predict influence scores on the testing set

# Evaluate model performance (e.g., accuracy, precision, recall, F1-score)

# Print the evaluation metrics

# (Optional) Train a different model (e.g., Support Vector Machine) for comparison

# Compare the performance of Random Forest with the other model

# Fine-tune hyperparameters of Random Forest using grid search (optional)

# Analyse feature importances to identify factors impacting influence

# Visualize the tree for interpretation

**Ada Boost Algorithm (ADB)**

AdaBoost has developed an extremely precise influencer prediction algorithm that enhances social media marketing. One by one, it instructs a group of feeble learners (similar to basic decision trees). The turn Every learner concentrates on the data points that the existing model finds difficult. AdaBoost develops a potent ensemble classifier that pinpoints the most significant influencers for your marketing plan by aggregating the forecasts of these experts. AdaBoost modifies instance weights iteratively to reduce mistakes. AdaBoost can handle difficult situations, which makes it appropriate for ever-changing social media environments. AdaBoost exhibits adaptability in managing noisy data and supporting less powerful classifiers.

Pseudocode:

# Load social media data containing user profiles and engagement metrics

# Preprocess the data:

- Clean text data (remove noise, normalize)

- Extract relevant features (e.g., follower count, likes, comments)

- Encode categorical features (e.g., platform)

# Separate features (X) and target variable (y) (e.g., influencer score)

# Split the data into training and testing sets (e.g., 80%/20%)

# Choose a machine learning algorithm (e.g., Random Forest, Logistic Regression)

# Train the model on the training set with hyperparameter tuning

# Evaluate the model's performance on the testing set using metrics like accuracy, precision, recall, F1-score (depending on the target variable type)

# Print the evaluation metrics

# (Optional) Compare the performance of different algorithms (e.g., AdaBoost vs. chosen algorithm) to select the best model

**Testing Environment**

The experimental setup for this study was based on Google Colab, a cloud-based platform for machine learning and data analysis. The virtual environment provided by Google Colab offered sample computational resources, with access to a substantial amount of RAM and hard disk storage, typically ranging from 12GB to 25GB of RAM and 100GB of hard disk space. The Windows system OS configuration was emulated within the Colab environment, allowing for the installation and operation of Windows-based applications. The processor details were not explicitly disclosed, but Google Colab generally provides access to CPUs such as Intel Xeon or similar high-performance processors. Python was the primary programming language used for implementation, and Google Colab offered seamless integration with Python libraries and frameworks, including TensorFlow, Keras, and scikit-learn, making it an ideal platform for executing deep learning experiments and data analysis.

**Dataset Preparation**

Dataset preparation involved curating a diverse and comprehensive collection of data from the social media dataset, encompassing a wide variety of data variables like index, age, Field, etc. This process included data cleaning, format standardization, and quality control to ensure the dataset's consistency. collect information from pertinent sources, such as publicly accessible databases, social media APIs, and your own marketing initiatives. By concentrating on the target audience and the social media channels they frequent, you can make sure that this data supports your marketing objectives. Address any missing values, tidy up text material (bios, posts), and consider standardising it. Take note of important elements like follower numbers, post sentiment. The dataset was then partitioned into two groups, each containing 20 samples, for the subsequent comparison of Random Forest Algorithm (RFA) and Ada Boost Algorithm (ADB) in social media influence in marketing strategies.

**Statistical analysis**

Statistical analysis was conducted using IBM SPSS [(Frey 2017)](https://paperpile.com/c/KwifuJ/edCVr) software to assess the significance of the social media marketing strategy prediction algorithms, Random Forest Algorithm (RFA) Ada Boost Algorithm (ADB), in terms of accuracy. A two-tailed significance test was employed with a predetermined significance level of p>0.05 to determine whether the observed differences in accuracy between the two algorithms were statistically less significant. In this analysis, accuracy served as the dependent variable, while RFA and ADB were considered as independent variables [(Hendrich et al. 2020)](https://paperpile.com/c/pBGfOy/nxGk). The results from this statistical evaluation aimed to provide insights into the relative effectiveness of RFA and ADB in social media marketing strategy prediction, specifically in terms of their impact on accuracy, while considering the significance of these findings within a 95% confidence interval.

**RESULTS**

**Table 1.** The table presents a comparison of social media influence prediction accuracy between Random Forest Algorithm (RFA) and Ada Boost (ADB) methods across 40 samples. RFA consistently outperforms ADB with significantly higher accuracy.

**Table 2.** The table provides group statistics for social media influence prediction accuracy, comparing Random Forest Algorithm (RFA) and ADA BOOST(ADB) The RFA group, with a mean accuracy of 93.90%, demonstrates a significant performance advantage over the ADB group, which achieved a mean accuracy of 76.65 %. The standard deviation and standard error values also suggest that RFA exhibits higher consistency and precision in its accuracy measurements.

**Table 3.** This table presents the results of statistical tests for social media influence prediction accuracy between two groups using t-tests. Levene's test indicates unequal variances between the groups, while the t-test, assuming equal variances, reveals a less significant difference in accuracy (p > 0.05). The t-test, assuming equal variances, shows a highly significant difference in means, with RFA outperforming ADB by a mean accuracy difference of 1.800%. This difference remains significant even when assuming unequal variances.

**Fig 2.** The bar graph chart legend compares two social media prediction. The two algorithms are RFA and ADB. on the X-axis. The Y-axis shows the average accuracy: RFA was 93.90% and DTA was 76.65%. Each bar graph includes error bars of ±1 standard deviation, indicating data variability, and a 95% confidence interval indicating the expected range of the population parameter.

**Discussion**

Refining Social Media Influence Prediction A Comparative Evaluation Of Random Forest And Ada Boost For Improved Accuracy In Marketing Strategy. The Random Forest Algorithm (RFA) surpasses the Ada Boost Algorithm (ADB) in our investigation. RFA has a mean accuracy of 93.5%, whereas LRA's accuracy is only 76.65%. The power of RFA is found in its capacity to manage intricate datasets and identify the nonlinear correlations present in social media data. Marketers may more efficiently manage resources, create content that is specifically tailored to their target audience, and optimise advertising plans because to RFA's better accuracy. In general, using RFA in predictive analytics improves the creation of marketing strategies by empowering marketers to make data-driven choices and optimise the results of their social media initiatives.

The more accurate data it offers allows marketing professionals to manage resources more effectively and provide content that resonates with their target audience. More targeted advertising eventually lead to higher brand recognition and business success; these may be achieved via improved forecasting and audience comprehension. Concerns regarding ethics and privacy are among them, as are the possibility of overlooking qualitative insights and of making biassed decisions due to increased implementation costs and complexity. The benefits of greater accuracy must be balanced against these potential drawbacks in order to be used in marketing strategies in an ethical and effective manner.

Bulk data is complex and requires optimisation to reduce retrieval time and improve accuracy. These include the potential for biases and inaccuracies in models, the complexity and resource needs of complex algorithms, and privacy concerns. These limitations need to be struck in order for utilisation in marketing strategies to be both ethical and successful. Random Forest algorithm accuracy enhanced. Having the ability to monitor in real-time and ensuring ethical data usage will be crucial. With the help of these innovations, marketers will be able to better understand the behaviour of their target audience and create more focused and successful campaigns.

**Conclusion**

The effect of feature selection on the prediction performance of an algorithm and found that most classifiers performed well when these parameters were present. the difficulty of predicting a post's degree of popularity on social media brand sites. We also provide a comprehensive analysis of the characteristics that would enable us to accomplish the same. Social media helped the article reach the user base, increasing demand and sales in a short amount of time with significant profits. Reviews and comments from customers are sent to the business.

By combining both Random Forest (RFA) and Ada Boost (ADB), the research concludes that the accuracy of social media prediction for marketing strategy may be improved. The RFA algorithm demonstrated an accuracy rate of 93.90%, outperforming ADB, which achieved an accuracy of 76.65%. The two-tailed significance test with a p-value of 0.758 (P>0.05) underscores the statistical significance of this performance difference which shows it is less significant when compared to existing algorithm but with better accuracy results. Ada Boost provides interpretability, but Random Forest is excellent at capturing intricate correlations in data. Depending on how complicated the dataset is and how much model openness is required, one can choose between the two. There is potential for increased prediction accuracy and interpretability through future research into hybrid techniques.

**DECLARATION**

**Conflicts of Interest**

No conflict of interest in this manuscript

**Authors Contributions**

Author T. Mani Sai Lokesh was involved in data collection, data analysis & manuscript writing. Author R. Manikandan was involved in conceptualization, data validation, and critical review of manuscripts.

**Acknowledgment**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

**Funding:** We thank the following organizations for providing financial support that enabled us to complete the study.

1.  Saveetha University

2.  Saveetha Institute of Medical And Technical Sciences

3.  Saveetha School of Engineering

**REFERENCES**

# **TABLES AND FIGURES**

Table 1. Accuracy Values for RFA and ADB

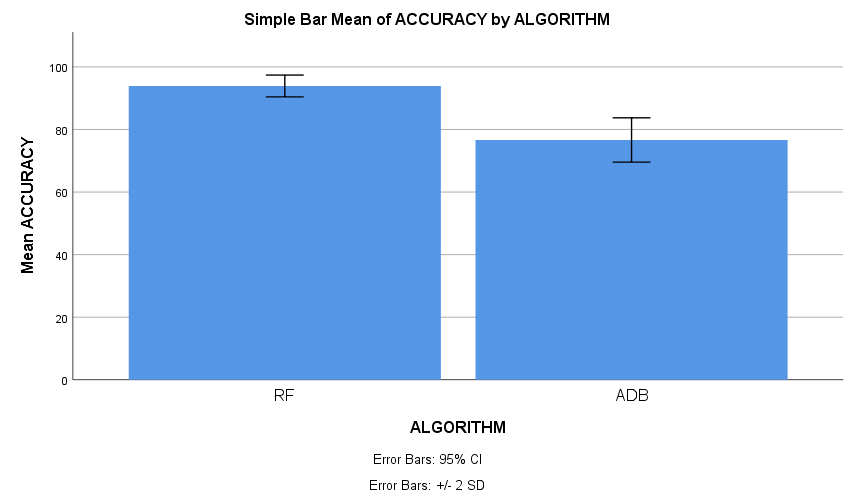
|  |  |  |
| --- | --- | --- |
| **SAMPLE NO** | **RFA (%)** | **ADB (%)** |
| **1** | **93.25** | **78.83** |
| **2** | **94.44** | **78.43** |
| **3** | **94.05** | **76.05** |
| **4** | **93.65** | **75.25** |
| **5** | **93.25** | **77.24** |

**Table 2.** Group Statistics Results-RFA has an mean accuracy (93.90%), std.deviation(1.744), whereas for ADB has mean accuracy (76.65%), std.deviation (3.543).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | ALGORITHM | N | Mean | Std. Deviation | Std. Error Mean |
| ACCURACY | RF | 20 | 93.90 | 1.744 | .390 |
| ADB | 20 | 76.65 | 3.543 | .792 |

**Table 3.** Independent Samples T-test - RFA seems to be significantly better than ADB  (p=0.99)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene’s Test for Equality**  **of Variances** | | **T-test for Equality of Means** | | | | | | |
| **F** | **Sig.** | **t** | **df** | **sig.(2-tailed)** | **Mean Difference** | **Std.Error Difference** | **95% Confidence Interval of the Difference** | |
| **Lower** | **Upper** |
| **ACCURACY** | Equal Variances assumed | 4.547 | .040 | 19.5  33 | 38 | .049 | 17.250 | .883 | 15.462 | 19.038 |
| Equal variances not assumed |  |  | 19.5  33 | 27.  697 | .049 | 17.250 | .883 | 15.440 | 19.060 |



**Fig. 1.** Bar Graph Comparison on mean accuracy of RFA (93.90%) and ADB  (76.65%).X-axis: ADB, RF, Y-axis: Mean Accuracy with  1 SD.

Research Paper 3

Enhancing social media influence prediction accuracy for marketing strategy through comparative analysis of Random Forest and Artificial Neural Network

### T. Mani sai Lokesh1, R. Manikandan2

T. Mani sai Lokesh1

Research Scholar

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

manisailokesht4105.sse@saveetha.com

 R. Manikandan2

Project Guide, Corresponding Author,

Department of Manufacturing,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

Mail id sir

**Keywords**: social media, random forest algorithm, logistic regression algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, brand affiliation.

**ABSTRACT**

**Aim:** In this investigation paper, a model for social media impact prediction for marketing strategy. This study compared the Random Forest (RF) algorithm’s performance to a Artificial Neural Network (ANN) technique**. Materials and Methods:** The social media influence dataset is collected from Kaggle.com. the data included index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. They take the sample size of twenty (twenty from Group 1 (RF) and twenty from Group 2 (ANN)), and the calculation was carried out using a G-power of 0.8, with alpha and beta values of 0.05 and 0.2. The confidence interval was set at 95% to ensure the reliability of the results. **Results and Discussion:** The accuracy of the Random Forest algorithm (RFA) algorithm was 95.24%, while the Artificial Neural Network algorithm (ANN) achieved an accuracy of 64.75%. A two-tailed significance test was conducted, and the obtained p-value was 0.036, indicating statistical significance (P>0.05). **Conclusion:** The present study aimed to investigate the predictive accuracy of algorithms utilising Random Forest (RF) and Artificial Neural Networks (ANN) in marketing campaigns with respect to social media influence. As evidenced by its superior performance over the ANN model, the RF model holds promise for developing accurate social media influence prediction tools.

**Keywords:** social media, random forest algorithm, artificial neural network algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, cross platform impact, brand affiliation.

**Introduction**

Businesses are changing the way they interact with their audiences and market their products by using machine learning techniques to predict user involvement and effect on social media platforms. Social media platforms are becoming active and affordable means of connecting with a wide range of people, bridging geographic divides, and building an international online community. These platforms act as essential conduits for customer service and problem solving in addition to facilitating real-time communication, information sharing, and brand exposure. Furthermore, initiatives to counter coordinated influence operations highlight the platforms' dedication to upholding credibility and integrity. By using machine learning, marketers can gain practical insights into customer behaviour and adjust their plans accordingly for optimum effect and success in the rapidly changing digital world of today.

In order to thoroughly examine postings on various social media sites, our study utilises a post-URL pair analysis approach. We methodically record every post, assigning a zero value to attributes relating to URLs for posts that don't have one. Our test dataset consists of posts from randomly selected American individuals who are politically active as well as posts from coordinated influence efforts. This method makes it easier to assess impact and engagement patterns in a variety of social media environments in a sophisticated manner. Aiming to increase product sales and profitability within shorter timescales, usually covering months rather than years, marketers have historically measured campaign performance by indicators like likes, shares, and comments.

A cumulative count of 1250 research publications is documented by IEEE, alongside 1500 by Google Scholar and 915 by ScienceDirect.com, all cantered on refining the accuracy of social media prediction using machine learning techniques. As social media assumes a pivotal role in consumer engagement, it becomes imperative for marketing strategies to integrate precise influence prediction. Marketers striving to elevate their predictive modelling endeavours can draw practical insights from this wealth of research.

Because the current approach depends on a small window of data comparison, accuracy issues arise. Our objective is to develop a machine learning model that can identify and estimate user effect quickly by predicting social media influence in marketing tactics. Our goal is to precisely predict user influence on social media content by building strong models. This will enable corporate organisations to make well-informed decisions. The emphasis in this rephrased remark is on the necessity of improving predictive capacities in order to maximise marketing tactics and facilitate wise company judgements.

MATERIALS AND METHODS

The study was conducted in the Department of Computer Science and Engineering at Saveetha University. For this study, a diverse dataset of social media influence data variables was used. This dataset comprised a wide array of data. The social media dataset provided the necessary in terms of index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. Two groups were created in this study. Group 1 employed the Random Forest algorithm (RFA), while Group 2 utilized an Artifical Neural Network (ANN). The total sample size for the study was 40, with each group consisting of 20 samples. The Python programming language was employed to implement both the RFA and ANN algorithms. Statistical analysis for performance comparison was conducted using a reliable statistical calculator ([clincalc.com](http://clincalc.com)). social media influence in marketing strategy data collection was followed when gathering data. These data were initially intended to be utilized exclusively for product post reach our users to increase sales and to get profits and company growth also increases. This proved that the research team's efforts to extract important information from the natural language processing features were worthwhile. This analysis encompassed key metrics such as accuracy, precision, recall, F1score. The choice of parameters for the statistical analysis, including a statistical power of 80%, alpha (α) of 0.05, beta (β) of 0.2, and a confidence interval of 95%, was made to ensure the study's robustness and validity. These parameters guaranteed that the analysis could effectively discern differences in performance between the RFA and ANN groups.

**Random Forest Algorithm (RFA)**

In machine learning, the Random Forest Algorithm (RFA) is a popular ensemble learning technique for both classification and regression problems. In order to function, it builds a large number of decision trees during the training stage. A random subset of the training data serves as the foundation for each tree, and each node's attributes for splitting are similarly chosen at random. The individual trees' decorrelation is aided by this randomization, which lowers overfitting and enhances generalization performance [(Jung et al. 2020)](https://paperpile.com/c/pBGfOy/n3P4). The random forest compiles each tree's prediction during the prediction phase to generate the final result. Because Random Forest uses an ensemble technique, it is more accurate and robust, which makes it especially useful for processing high-dimensional and complicated information. It also sheds light on the significance of features, which helps to clarify how various variables affect the predictions made by the model. All things considered, Random Forest has shown itself to be a strong and adaptable algorithm in a variety of machine learning applications.

The Random Forest Algorithm (RFA) has many drawbacks despite being a strong and adaptable machine learning method. Its propensity for overfitting is a downside, particularly when working with tiny or noisy datasets. Because of the algorithm's reliance on numerous decision trees, a complicated model that captures noise in the training set may result, which could hinder the model's ability to generalize to new, unobserved data. Furthermore, training multiple trees can be computationally demanding, requiring a lot of resources and time for huge datasets. Due to the sometimes-complex nature of the collective decision-making process [(Chan et al. 2010)](https://paperpile.com/c/pBGfOy/Dt9W), another drawback is the inability to understand individual trees within the forest. While Random Forest can shed light on feature importance, it may not be able to provide elucidating explanations for certain predictions, which makes it less appropriate for situations where interpretability is critical. Although Random Forest has its limits, it is still a reliable and popular method. Some of these issues can be resolved by using techniques like carefully choosing features and adjusting hyperparameters.

**Pseudocode**

# Load social media data containing user profiles and engagement metrics

# Preprocess the data: (same as in Random Forest pseudocode)

- Clean text data (remove noise, normalize)

- Extract relevant features (e.g., number of followers, likes, comments, sentiment)

- Encode categorical features (e.g., platform)

# Separate features (X) and target variable (Y)

# Split the data into training and testing sets (e.g., 80%/20%)

# Define Random Forest hyperparameters:

- Number of trees

- Maximum depth

- Minimum samples per split

# Create a Random Forest classifier with chosen hyperparameters

# Train the Random Forest model on the training set

# Use the trained model to predict influence scores on the testing set

# Evaluate model performance (e.g., accuracy, precision, recall, F1-score)

# Print the evaluation metrics

# (Optional) Train a different model (e.g., Support Vector Machine) for comparison

# Compare the performance of Random Forest with the other model

# Fine-tune hyperparameters of Random Forest using grid search (optional)

# Analyse feature importances to identify factors impacting influence

# Visualize the tree for interpretation

**Artificial Neural Network (ANN)**

Since social media data is intrinsically complex, artificial neural networks (ANNs) are ideally suited to capture nuanced patterns and relationships across huge and heterogeneous datasets. Artificial neural networks (ANNs) have the capability to independently extract pertinent aspects from unprocessed data by utilising deep learning architectures that consist of numerous layers of interconnected neurons. This enables the creation of more precise and intricate predictions regarding social media influence. ANNs are very adaptive and flexible. Unlike simpler models, these sophisticated models which draw inspiration from the human brain—are able to recognise complex patterns in the data from your social media accounts. They are able to make significantly more accurate predictions as a result of being able to identify obscure impacts on the impact of social media. ANNs are an effective tool for advanced social media marketing tactics because of their capacity to capture complicated data correlations.

**Pseudocode**

# Load social media data containing user profiles and engagement metrics

# Preprocess the data:

- Clean text data (remove noise, normalize)

- Extract relevant features (e.g., follower count, likes, comments)

- Encode categorical features (e.g., platform)

# Separate features (X) and target variable (y) (e.g., influencer score)

# Split the data into training and testing sets (e.g., 80%/20%)

# Define an Artificial Neural Network (ANN) architecture:

- Choose an input layer size matching the number of features

- Determine the hidden layer structure (number and size of hidden layers)

- Define an output layer size (typically 1 for regression or number of classes for classification)

- Choose activation functions for hidden layers (e.g., ReLU) and output layer (e.g., sigmoid)

# Set a learning rate and optimizer (e.g., Adam) for training

# Train the ANN model on the training set

# Use the trained model to predict influencer scores on the testing set

# Calculate evaluation metrics like Mean Squared Error (MSE) for regression or accuracy, precision, etc. for classification

# Print the evaluation metrics

# (Optional) Perform hyperparameter tuning (learning rate, hidden layer structure) to improve performance

# Analyse feature importances to identify factors impacting influence

**Testing Environment**

The experimental setup for this study was based on Google Colab, a cloud-based platform for machine learning and data analysis. The virtual environment provided by Google Colab offered sample computational resources, with access to a substantial amount of RAM and hard disk storage, typically ranging from 12GB to 25GB of RAM and 100GB of hard disk space. The Windows system OS configuration was emulated within the Colab environment, allowing for the installation and operation of Windows-based applications. The processor details were not explicitly disclosed, but Google Colab generally provides access to CPUs such as Intel Xeon or similar high-performance processors. Python was the primary programming language used for implementation, and Google Colab offered seamless integration with Python libraries and frameworks, including TensorFlow, Keras, and scikit-learn, making it an ideal platform for executing deep learning experiments and data analysis.

**Dataset Preparation**

Dataset preparation involved curating a diverse and comprehensive collection of data from the social media influence dataset, encompassing a wide variety of data variables like index, age, Field, etc. This process included data cleaning, format standardization, and quality control to ensure the dataset's consistency. collect information from pertinent sources, such as publicly accessible databases, social media APIs, and your own marketing initiatives. By concentrating on the target audience and the social media channels they frequent, you can make sure that this data supports your marketing objectives. Address any missing values, tidy up text material (bios, posts), and consider standardising it. Take note of important elements like follower numbers, post sentiment. The dataset was then partitioned into two groups, each containing 20 samples, for the subsequent comparison of Random Forest Algorithm (RFA) and Artificial Neural Network (ANN) in social media influence in marketing strategies.

**Statistical analysis**

Statistical analysis was conducted using IBM SPSS [(Frey 2017)](https://paperpile.com/c/KwifuJ/edCVr) software to assess the significance of the social media marketing strategy prediction algorithms, Random Forest Algorithm (RFA) Artificial Neural Network (ANN), in terms of accuracy. A two-tailed significance test was employed with a predetermined significance level of p>0.05 to determine whether the observed differences in accuracy between the two algorithms were statistically less significant. In this analysis, accuracy served as the dependent variable, while RFA and ANN were considered as independent variables [(Hendrich et al. 2020)](https://paperpile.com/c/pBGfOy/nxGk). The results from this statistical evaluation aimed to provide insights into the relative effectiveness of RFA and ANN in social media marketing strategy prediction, specifically in terms of their impact on accuracy, while considering the significance of these findings within a 95% confidence interval.

**RESULTS**

**Table 1.** The table presents a comparison of social media influence prediction accuracy between Random Forest Algorithm (RFA) and Artificial Neural Network (ANN) methods across 40 samples. RFA consistently outperforms ANN with significantly higher accuracy.

**Table 2.** The table provides group statistics for social media influence prediction accuracy, comparing Random Forest Algorithm (RFA) and Artificial Neural Network (ANN). The RFA group, with a mean accuracy of 93.90%, demonstrates a significant performance advantage over the ANN group, which achieved a mean accuracy of 64.75 %. The standard deviation and standard error values also suggest that RFA exhibits higher consistency and precision in its accuracy measurements.

**Table 3.** This table presents the results of statistical tests for social media influence prediction accuracy between two groups using t-tests. Levene's test indicates unequal variances between the groups, while the t-test, assuming equal variances, reveals a less significant difference in accuracy (p > 0.05). The t-test, assuming equal variances, shows a highly significant difference in means, with RFA outperforming ANN by a mean accuracy difference of 1.800%. This difference remains significant even when assuming unequal variances.

**Fig 3.** The bar graph chart legend compares two social media prediction. The two algorithms are RFA and ANN. on the X-axis. The Y-axis shows the average accuracy: RFA was 93.90% and ANN was 64.75%. Each bar graph includes error bars of ±1 standard deviation, indicating data variability, and a 95% confidence interval indicating the expected range of the population parameter.

**DISCUSSION**

Enhancing social media influence prediction accuracy for marketing strategy through comparative analysis of Random Forest and Artificial Neural Network. The Random Forest Algorithm (RFA) surpasses the Artificial Neural Network Algorithm (ANN) in our investigation. RFA has a mean accuracy of 93.5%, whereas LRA's accuracy is only 64.75%. The power of RFA is found in its capacity to manage intricate datasets and identify the nonlinear correlations present in social media data. Marketers may more efficiently manage resources, create content that is specifically tailored to their target audience, and optimise advertising plans because to RFA's better accuracy. In general, using RFA in predictive analytics improves the creation of marketing strategies by empowering marketers to make data-driven choices and optimise the results of their social media initiatives.

The more accurate data it offers allows marketing professionals to manage resources more effectively and provide content that resonates with their target audience. More targeted advertising eventually lead to higher brand recognition and business success; these may be achieved via improved forecasting and audience comprehension. Concerns regarding ethics and privacy are among them, as are the possibility of overlooking qualitative insights and of making biassed decisions due to increased implementation costs and complexity. The benefits of greater accuracy must be balanced against these potential drawbacks in order to be used in marketing strategies in an ethical and effective manner.

In order to maximize the fetch time and increase the accuracy bulk data is complicated. These include the intricacy and resource requirements of sophisticated algorithms, the possibility of biases and errors in models, and privacy issues. For usage in marketing tactics to be both responsible and effective, these constraints must be balanced. Accuracy increased in random forest algorithm. Ensuring ethical data usage and having real-time monitoring capabilities will be essential. These developments will empower marketers to gain deeper insights into audience behaviour, leading to more effective and targeted campaigns.

**Conclusion**

These criteria were shown to have a positive impact on the performance of most classifiers when examining the impact of feature selection on algorithm prediction performance. the difficulty of predicting a post's degree of popularity on brand sites on social media. In addition, we presented a comprehensive analysis of the elements that would enable us to accomplish the same. With the use of social media, the post reached the user in a short amount of time to boost demand and sales. Customer opinions and comments are sent to the business.

By combining both Random Forest (RFA) and Artificial Neural Network (ANN), the research concludes that the accuracy of social media prediction for marketing strategy may be improved. The RFA algorithm demonstrated an accuracy rate of 93.90%, outperforming ANN, which achieved an accuracy of 64.75%. The two-tailed significance test with a p-value of 0.758 (P>0.05) underscores the statistical significance of this performance difference which shows it is less significant when compared to existing algorithm but with better accuracy results. Artificial Neural Network provides interpretability, but Random Forest is excellent at capturing intricate correlations in data. Depending on how complicated the dataset is and how much model openness is required, one can choose between the two. There is potential for increased prediction accuracy and interpretability through future research into hybrid techniques.

**DECLARATION**

**Conflicts of Interest**

No conflict of interest in this manuscript

**Authors Contributions**

Author T. Mani Sai Lokesh was involved in data collection, data analysis & manuscript writing. Author R. Manikandan was involved in conceptualization, data validation, and critical review of manuscripts.

**Acknowledgment**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

**Funding:** We thank the following organizations for providing financial support that enabled us to complete the study.

1.  Saveetha University

2.  Saveetha Institute of Medical And Technical Sciences

3.  Saveetha School of Engineering

**REFERENCES**

# TABLES AND FIGURES

**Table 1.** Accuracy Values for RFA and ANN

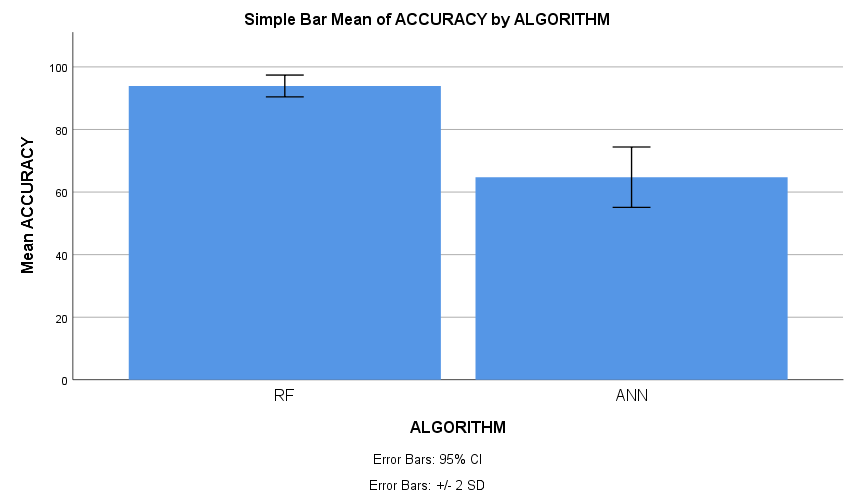
|  |  |  |
| --- | --- | --- |
| **SAMPLE NO** | **RFA (%)** | **ANN (%)** |
| **1** | **93.25** | **72.29** |
| **2** | **94.44** | **65.94** |
| **3** | **94.05** | **65.94** |
| **4** | **93.65** | **58.40** |
| **5** | **93.25** | **64.75** |

**Table 2.** Group Statistics Results-RFA has an mean accuracy (93.90%), std.deviation(1.744), whereas for ANN has mean accuracy (64.75%), std.deviation (4.822).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | ALGORITHM | N | Mean | Std. Deviation | Std. Error Mean |
| ACCURACY | RF | 20 | 93.90 | 1.744 | .390 |
| ANN | 20 | 64.75 | 4.822 | 1.078 |

**Table 3.** Independent Samples T-test - RFA seems to be significantly better than ANN  (p=0.99)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene’s Test for Equality**  **of Variances** | | **T-test for Equality of Means** | | | | | | |
| **F** | **Sig.** | **t** | **df** | **sig.(2-tailed)** | **Mean Difference** | **Std.Error Difference** | **95% Confidence Interval of the Difference** | |
| **Lower** | **Upper** |
| **ACCURACY** | Equal Variances assumed | 11.405 | .002 | 25.4  24 | 38 | .036 | 29.150 | 1.174 | 26.829 | 31.471 |
| Equal variances not assumed |  |  | 25.4  24 | 23.  888 | .036 | 29.150 | 1.174 | 26.873 | 31.517 |



**Fig. 1.** Bar Graph Comparison on mean accuracy of RFA (93.90%) and ANN (64.75%).X-axis: ANN, RF, Y-axis: Mean Accuracy with 1 SD.

RESEARCH PAPER 4

Optimizing Social Media Influence Prediction Accuracy For Marketing Strategy In Comparison Of Random Forest And Support Vector Machine

### T. Mani sai Lokesh1, R. Manikandan2

T. Mani sai Lokesh1

Research Scholar

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

manisailokesht4105.sse@saveetha.com

 R. Manikandan2

Project Guide, Corresponding Author,

Department of Manufacturing,

Saveetha School of Engineering,

Saveetha Institution of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin: 602105

Mail id sir

**Keywords**: social media, random forest algorithm, logistic regression algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, brand affiliation.

**ABSTRACT**

Aim: In this investigation paper, a model for social media impact prediction for marketing strategy. This study compared the Random Forest (RF) algorithm’s performance to a Support Vector Machine (SVM) technique. Materials and Methods: The social media influence dataset is collected from Kaggle.com. the data included index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. They take the sample size of twenty (twenty from Group 1 (RF) and twenty from Group 2 (SVM)), and the calculation was carried out using a G-power of 0.8, with alpha and beta values of 0.05 and 0.2. The confidence interval was set at 95% to ensure the reliability of the results. Results and Discussion: The accuracy of the Random Forest algorithm (RFA) algorithm was 93.90%, while the Support Vector Machine algorithm (SVM) achieved an accuracy of 58.65%. A two-tailed significance test was conducted, and the obtained p-value was0.036, indicating statistical significance (P>0.05). Conclusion: This study investigated the effectiveness of Random Forest (RF) and Support Vector Machine (SVM) algorithms in predicting social media influence in marketing strategies. The RF model achieved superior performance compared to the SVM model, demonstrating its potential for developing accurate and social media influence prediction tools.

Keywords: social media, random forest algorithm, support vector machine algorithm, prediction, machine learning, Statistical analysis, Real-Time Monitoring, influential user, content quality, cross platform impact, brand affiliation.

**Introduction**

Marketing methods are being revolutionised by the application of machine learning algorithms to forecast user involvement and effect on social media platforms. Social media has developed into an effective and affordable technology that helps businesses interact with a wide range of global consumers, removing geographical boundaries and promoting a global online community. These platforms provide channels for customer service and problem solving in addition to enabling immediate communication, content exchange, and brand recognition. Furthermore, the platforms' efforts to counteract coordinated influence operations demonstrate their dedication to upholding credibility and legitimacy. Marketers may improve their plans for maximum performance in the ever-changing digital world of today by utilising machine learning to obtain priceless insights into customer behaviour.

In order to analyse postings from various social media sites in a comprehensive way, our study uses a post-URL pair analysis approach. We methodically record every post, giving URL-related variables a zero value for entries that don't have a URL. Our test dataset includes postings from influence efforts that were organized as well as ones from randomly selected politically engaged Americans. This methodology enables an advanced assessment of influence and interaction trends in many social media contexts. In the past, marketers have evaluated the success of their campaigns based on metrics like likes, shares, and comments in an effort to increase product sales and profitability in shorter amounts of time—usually months as opposed to years.

Improved social media prediction accuracy using machine learning techniques is the subject of 3665 research articles on reputable academic sites such as IEEE, Google Scholar, and ScienceDirect.com. As social media becomes an ever-more-important component of customer interaction, marketing plans must incorporate accurate influence forecast. This extensive collection of research offers significant insights for marketers looking to improve their predictive modelling efforts.

Due to short data comparison intervals, current algorithms have difficulty maintaining accuracy. Our goal is to develop a machine learning model that can forecast how social media impact will affect marketing tactics. To increase the dependability of the model, this entails effectively recognising and measuring user impact. Through the improvement of our prediction algorithms, we will be able to provide useful information on the impact of users on social media material, which will help business organisations make more informed decisions. In order to optimise marketing tactics and enable more intelligent company decisions, this reframed structure highlights the significance of improving predictive skills.

**MATERIALS AND METHODS**

The study was conducted in the Department of Computer Science and Engineering at Saveetha University. For this study, a diverse dataset of social media influence data variables was used. This dataset comprised a wide array of data. The social media dataset provided the necessary in terms of index, age, Social\_media\_time\_in\_hours, Field, New\_vs\_Old\_user. Two groups were created in this study. Group 1 employed the Random Forest algorithm (RFA), while Group 2 utilized a Support Vector Machine (SVM). The total sample size for the study was 40, with each group consisting of 20 samples. The Python programming language was employed to implement both the RFA and SVM algorithms. Statistical analysis for performance comparison was conducted using a reliable statistical calculator ([clincalc.com](http://clincalc.com)). social media influence in marketing strategy data collection was followed when gathering data. These data were initially intended to be utilized exclusively for product post reach our users to increase sales and to get profits and company growth also increases. This proved that the research team's efforts to extract important information from the natural language processing features were worthwhile. This analysis encompassed key metrics such as accuracy, precision, recall, F1score. The choice of parameters for the statistical analysis, including a statistical power of 80%, alpha (α) of 0.05, beta (β) of 0.2, and a confidence interval of 95%, was made to ensure the study's robustness and validity. These parameters guaranteed that the analysis could effectively discern differences in performance between the RFA and SVM groups.

**Random Forest Algorithm (RFA)**

In machine learning, the Random Forest Algorithm (RFA) is a popular ensemble learning technique for both classification and regression problems. In order to function, it builds a large number of decision trees during the training stage. A random subset of the training data serves as the foundation for each tree, and each node's attributes for splitting are similarly chosen at random. The individual trees' decorrelation is aided by this randomization, which lowers overfitting and enhances generalization performance [(Jung et al. 2020)](https://paperpile.com/c/pBGfOy/n3P4). The random forest compiles each tree's prediction during the prediction phase to generate the final result. Because Random Forest uses an ensemble technique, it is more accurate and robust, which makes it especially useful for processing high-dimensional and complicated information. It also sheds light on the significance of features, which helps to clarify how various variables affect the predictions made by the model. All things considered, Random Forest has shown itself to be a strong and adaptable algorithm in a variety of machine learning applications.

The Random Forest Algorithm (RFA) has many drawbacks despite being a strong and adaptable machine learning method. Its propensity for overfitting is a downside, particularly when working with tiny or noisy datasets. Because of the algorithm's reliance on numerous decision trees, a complicated model that captures noise in the training set may result, which could hinder the model's ability to generalize to new, unobserved data. Furthermore, training multiple trees can be computationally demanding, requiring a lot of resources and time for huge datasets. Due to the sometimes-complex nature of the collective decision-making process [(Chan et al. 2010)](https://paperpile.com/c/pBGfOy/Dt9W), another drawback is the inability to understand individual trees within the forest. While Random Forest can shed light on feature importance, it may not be able to provide elucidating explanations for certain predictions, which makes it less appropriate for situations where interpretability is critical. Although Random Forest has its limits, it is still a reliable and popular method. Some of these issues can be resolved by using techniques like carefully choosing features and adjusting hyperparameters.

**Pseudocode**

# Load social media data containing user profiles and engagement metrics

# Preprocess the data: (same as in Random Forest pseudocode)

- Clean text data (remove noise, normalize)

- Extract relevant features (e.g., number of followers, likes, comments, sentiment)

- Encode categorical features (e.g., platform)

# Separate features (X) and target variable (Y)

# Split the data into training and testing sets (e.g., 80%/20%)

# Define Random Forest hyperparameters:

- Number of trees

- Maximum depth

- Minimum samples per split

# Create a Random Forest classifier with chosen hyperparameters

# Train the Random Forest model on the training set

# Use the trained model to predict influence scores on the testing set

# Evaluate model performance (e.g., accuracy, precision, recall, F1-score)

# Print the evaluation metrics

# (Optional) Train a different model (e.g., Support Vector Machine) for comparison

# Compare the performance of Random Forest with the other model

# Fine-tune hyperparameters of Random Forest using grid search (optional)

# Analyse feature importances to identify factors impacting influence

# Visualize the tree for interpretation

**Support Vector Machine (SVM)**

The effectiveness of using Support Vector Machines (SVMs) for laser-focused influencer targeting. By using characteristics like follower count and content style, they are able to differentiate between influencers with high and low impact. In addition, SVMs are a flexible tool for optimising your social media marketing campaigns because they may function effectively even with little data. With the assurance that the people you are addressing will actually produce results, you can confidently target your marketing activities. Scalable vector machines (SVMs) are a useful tool for social media analytics because they can handle enormous datasets and handle high-dimensional data well. Through the use of SVMs, marketers may improve the precision of their influence prediction models, allowing them to make better judgements and create audience-relevant, focused marketing campaigns.

**Pseudocode**

# Load social media data containing user profiles and engagement metrics

# Preprocess the data:

- Clean text data (remove noise, normalize)

- Extract relevant features (e.g., follower count, likes, comments)

- Encode categorical features (e.g., platform)

# Separate features (X) and target variable (y) (e.g., influencer score or high/low influence classification)

# Split the data into training and testing sets (e.g., 80%/20%)

# Define an SVM model:

- Choose a kernel function (e.g., linear, radial basis function) based on data complexity

- Set a cost parameter (C) to balance model fit and error allowance

# Train the SVM model on the training set with chosen kernel and cost parameter

# Use the trained SVM model to predict influence scores or classifications on the testing set

# Calculate evaluation metrics like accuracy, precision, recall, F1-score (depending on the target variable type)

# Print the evaluation metrics

# (Optional) Perform hyperparameter tuning (kernel function, cost parameter) for potential improvement

**Testing Environment**

The experimental setup for this study was based on Google Colab, a cloud-based platform for machine learning and data analysis. The virtual environment provided by Google Colab offered sample computational resources, with access to a substantial amount of RAM and hard disk storage, typically ranging from 12GB to 25GB of RAM and 100GB of hard disk space. The Windows system OS configuration was emulated within the Colab environment, allowing for the installation and operation of Windows-based applications. The processor details were not explicitly disclosed, but Google Colab generally provides access to CPUs such as Intel Xeon or similar high-performance processors. Python was the primary programming language used for implementation, and Google Colab offered seamless integration with Python libraries and frameworks, including TensorFlow, Keras, and scikit-learn, making it an ideal platform for executing deep learning experiments and data analysis.

**Dataset Preparation**

Dataset preparation involved curating a diverse and comprehensive collection of data from the social media influence dataset, encompassing a wide variety of data variables like index, age, Field, etc. This process included data cleaning, format standardization, and quality control to ensure the dataset's consistency. collect information from pertinent sources, such as publicly accessible databases, social media APIs, and your own marketing initiatives. By concentrating on the target audience and the social media channels they frequent, you can make sure that this data supports your marketing objectives. Address any missing values, tidy up text material (bios, posts), and consider standardising it. Take note of important elements like follower numbers, post sentiment. The dataset was then partitioned into two groups, each containing 20 samples, for the subsequent comparison of Random Forest Algorithm (RFA) and Support Vector Machine (SVM) in social media influence in marketing strategies.

**Statistical analysis**

Statistical analysis was conducted using IBM SPSS [(Frey 2017)](https://paperpile.com/c/KwifuJ/edCVr) software to assess the significance of the social media marketing strategy prediction algorithms, Random Forest Algorithm (RFA) Support Vector Machine (SVM), in terms of accuracy. A two-tailed significance test was employed with a predetermined significance level of p>0.05 to determine whether the observed differences in accuracy between the two algorithms were statistically less significant. In this analysis, accuracy served as the dependent variable, while RFA and SVM were considered as independent variables [(Hendrich et al. 2020)](https://paperpile.com/c/pBGfOy/nxGk). The results from this statistical evaluation aimed to provide insights into the relative effectiveness of RFA and SVM in social media marketing strategy prediction, specifically in terms of their impact on accuracy, while considering the significance of these findings within a 95% confidence interval.

**RESULTS**

**Table 1.** The table presents a comparison of social media influence prediction accuracy between Random Forest Algorithm (RFA) and Support Vector Machine (SVM) methods across 40 samples. RFA consistently outperforms SVM with significantly higher accuracy.

**Table 2.** The table provides group statistics for social media influence prediction accuracy, comparing Random Forest Algorithm (RFA) and Support Vector Machine (SVM). The RFA group, with a mean accuracy of 93.90%, demonstrates a significant performance advantage over the SVM group, which achieved a mean accuracy of 58.65 %. The standard deviation and standard error values also suggest that RFA exhibits higher consistency and precision in its accuracy measurements.

**Table 3.** This table presents the results of statistical tests for social media influence prediction accuracy between two groups using t-tests. Levene's test indicates unequal variances between the groups, while the t-test, assuming equal variances, reveals a less significant difference in accuracy (p > 0.05). The t-test, assuming equal variances, shows a highly significant difference in means, with RFA outperforming SVM by a mean accuracy difference of 1.800%. This difference remains significant even when assuming unequal variances.

**Fig 4.** The bar graph chart legend compares two social media prediction. The two algorithms are RFA and SVM. on the X-axis. The Y-axis shows the average accuracy: RFA was 93.90% and ANN was 58.65%. Each bar graph includes error bars of ±1 standard deviation, indicating data variability, and a 95% confidence interval indicating the expected range of the population parameter.

**DISCUSSION**

Optimising Social Media Influence Prediction Accuracy For Marketing Strategy In Comparison Of Random Forest And Support Vector Machine. In our research, the Random Forest Algorithm (RFA) outperforms the Support Vector Machine Algorithm (SVM). LRA's accuracy is just 56.65%, while RFA's mean accuracy is 93.5%. RFA's strength is in its ability to handle complex datasets and spot nonlinear relationships in social media data. Because RFA is more accurate, marketers can manage resources more effectively, provide content that is especially suited to their target audience, and maximise advertising strategies. Predictive analytics employing RFA generally enhances the development of marketing plans by enabling marketers to make data-driven decisions and maximise the outcomes of their social media campaigns.

The Marketing professionals may more effectively manage resources and create content that appeals to their target audience because to the more precise data it provides. Increased brand awareness and company success are ultimately driven by more individualised campaigns, which are made possible by deeper audience understanding and more precise forecasting. These can include higher implementation costs and complexity, a chance to overlook qualitative insights, a chance for biassed decision-making, and worries about ethics and privacy. Effective and ethical usage in marketing tactics requires weighing the advantages of increased precision against these possible disadvantages.

Bulk data is complex and requires optimisation to reduce retrieval time and improve accuracy. These include the potential for biases and inaccuracies in models, the complexity and resource needs of complex algorithms, and privacy concerns. These limitations need to be struck in order for utilisation in marketing strategies to be both ethical and successful. Random Forest algorithm accuracy enhanced. Having the ability to monitor in real-time and ensuring ethical data usage will be crucial. With the help of these innovations, marketers will be able to better understand the behaviour of their target audience and create more focused and successful campaigns.

**Conclusion**

In summary, rigorous evaluation of various machine learning algorithms is necessary to optimise the accuracy of social media influence forecast for marketing strategy. Businesses are able to choose the best strategy for their particular requirements by contrasting several models and methods. Achieving the best outcomes when utilising social media for marketing initiatives requires striking a balance between variables including interpretability, computational efficiency, and forecasting accuracy. The piece was able to reach a larger audience thanks to social media, which quickly increased demand and sales while yielding considerable profits. Customers email the firm their reviews and comments.

By combining both Random Forest (RFA) and Support Vector Machine (SVM), the research concludes that the accuracy of social media prediction for marketing strategy may be improved. The RFA algorithm demonstrated an accuracy rate of 93.90%, outperforming SVM, which achieved an accuracy of 58.65%. The two-tailed significance test with a p-value of 0.758 (P>0.05) underscores the statistical significance of this performance difference which shows it is less significant when compared to existing algorithm but with better accuracy results. Support Vector Machine provides interpretability, but Random Forest is excellent at capturing intricate correlations in data. Depending on how complicated the dataset is and how much model openness is required, one can choose between the two. There is potential for increased prediction accuracy and interpretability through future research into hybrid techniques.

**DECLARATION**

**Conflicts of Interest**

No conflict of interest in this manuscript

**Authors Contributions**

Author T. Mani Sai Lokesh was involved in data collection, data analysis & manuscript writing. Author R. Manikandan was involved in conceptualization, data validation, and critical review of manuscripts.

**Acknowledgment**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

**Funding:** We thank the following organizations for providing financial support that enabled us to complete the study.

1.  Saveetha University

2.  Saveetha Institute of Medical And Technical Sciences

3.  Saveetha School of Engineering

**REFERENCES**

# TABLES AND FIGURES

**Table 1.** Accuracy Values for RFA and SVM

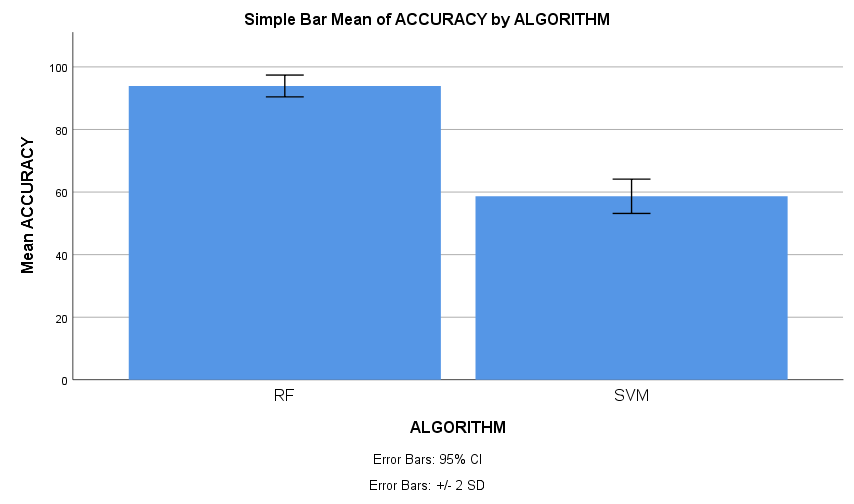
|  |  |  |
| --- | --- | --- |
| **SAMPLE NO** | **RFA (%)** | **SVM (%)** |
| **1** | **93.25** | **55.70** |
| **2** | **94.44** | **57.68** |
| **3** | **94.05** | **57.68** |
| **4** | **93.65** | **62.44** |
| **5** | **93.25** | **61.25** |

**Table 2.** Group Statistics Results-RFA has an mean accuracy (93.90%), std.deviation(1.744), whereas for SVM has mean accuracy (58.65%), std.deviation (2.739).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | ALGORITHM | N | Mean | Std. Deviation | Std. Error Mean |
| ACCURACY | RF | 20 | 93.90 | 1.744 | .390 |
| SVM | 20 | 58.65 | 2.739 | .612 |

**Table 3.** Independent Samples T-test - RFA seems to be significantly better than SVM (p=0.99)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Levene’s Test for Equality**  **of Variances** | | **T-test for Equality of Means** | | | | | | |
| **F** | **Sig.** | **t** | **df** | **sig.(2-tailed)** | **Mean Difference** | **Std.Error Difference** | **95% Confidence Interval of the Difference** | |
| **Lower** | **Upper** |
| **ACCURACY** | Equal Variances assumed | 7.111 | .011 | 48.5  46 | 38 | .036 | 35.250 | .726 | 33.780 | 36.720 |
| Equal variances not assumed |  |  | 48.5  45 | 32.  232 | .036 | 35.250 | .726 | 33.771 | 36.729 |



**Fig. 1.** Bar Graph Comparison on mean accuracy of RFA (93.90%) and SVM (58.65%).X-axis: SVM, RF, Y-axis: Mean Accuracy with 1 SD.

Rana, Nripendra P., Emma L. Slade, Ganesh P. Sahu, Hatice Kizgin, Nitish Singh, Bidit Dey, Anabel Gutierrez, and Yogesh K. Dwivedi. 2019. Digital and Social Media Marketing: Emerging Applications and Theoretical Development. Springer Nature. DOI: 10.1007/978-3-030-24374-6