SOCIAL MEDIA POPULARITY PREDICTION USING MACHINE LEARNING

1. **V. Ramya, Asst. prof CSE dept, Gokula Krishna College of Engineering, M. Tech (CSE), Sullurpet, Tirupati District, AP**
2. **K. Muni kumar, R. Bhanu Prasad, G. Devi, V. Dharanika ,K. Sudha , B.Tech UG Scholars, Gokula Krishna College of Engineering, Sullurpet, Tirupati District, AP**

ABSTRACT

This study presents an advanced machine learning–driven framework for predicting the popularity of social media content by integrating multi-dimensional features and adaptive learning models. Unlike traditional approaches that rely on limited statistical indicators or manual analysis, the proposed system leverages a combination of textual, contextual, and user-centric attributes to enhance prediction accuracy. The framework incorporates natural language processing techniques to extract semantic meaning from captions, hashtags, and descriptions, while also utilizing numerical features such as posting time, user engagement history, and interaction patterns.

To improve predictive performance, multiple machine learning algorithms, including Random Forest, Support Vector Machines, and deep learning models such as Convolutional Neural Networks and Long Short-Term Memory networks, are evaluated and optimized. The system dynamically learns feature importance and adapts to evolving social media trends, overcoming the limitations of static models. Additionally, sentiment analysis is integrated to capture audience perception, further strengthening prediction capability.

Experimental evaluation demonstrates that the proposed approach achieves superior accuracy and robustness compared to conventional methods. By providing data-driven insights and optimal content strategies, the model enables users and businesses to maximize reach, engagement, and visibility in competitive social media environments.

Keywords— Social media popularity prediction, machine learning, deep learning, natural language processing, feature extraction, sentiment analysis, multi-modal data, user engagement modeling, regression models, predictive analytics.

I. INTRODUCTION

Social media platforms have emerged as powerful communication channels that enable users to create, share, and interact with digital content at an unprecedented scale. With the exponential growth of platforms such as Flickr, Instagram, and Twitter, understanding and predicting the popularity of user-generated content has become a critical research problem. Popularity prediction refers to estimating the level of engagement a post will receive, typically measured through metrics such as views, likes, shares, and comments. Accurate prediction of content popularity has significant applications in targeted advertising, recommendation systems, trend analysis, and digital marketing strategies.

However, predicting social media popularity is inherently complex due to the dynamic and multifaceted nature of influencing factors. Content quality, user influence, posting time, contextual relevance, and audience behavior all contribute to popularity, making it difficult to model using traditional approaches. Early studies primarily focused on single-modality features such as image content or textual information [1], while others incorporated temporal dynamics to capture user interaction trends over time [2]. Additionally, user behavior and network structures were explored to understand social influence patterns [3].

Recent advancements have emphasized the importance of multi-modal learning, where different types of data such as images, text, and metadata are combined to improve prediction accuracy [4], [5]. These approaches demonstrate that integrating visual and textual information leads to more robust performance compared to single-feature models. However, many existing methods rely on simple feature concatenation, which fails to capture complex relationships between different modalities [7], [19].

With the evolution of machine learning and deep learning, more sophisticated techniques have been introduced. Deep neural networks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been applied to extract high-level representations from large-scale data [15], [13]. Furthermore, the introduction of attention mechanisms has significantly improved feature representation by enabling models to focus on the most relevant information within input data [9]. This has led to more effective handling of heterogeneous data sources.

Despite these advancements, several challenges remain unresolved. Many models struggle with high-dimensional data, scalability issues, and the inability to adapt to rapidly changing social media trends. Additionally, traditional machine learning approaches often lack the capability to dynamically weigh the importance of different features, leading to suboptimal performance.

To address these limitations, this research proposes a machine learning-based framework that integrates multi-modal features, including textual content, user information, and numerical attributes, for improved popularity prediction. By incorporating advanced learning techniques such as ensemble modeling and natural language processing, the proposed system aims to enhance prediction accuracy while maintaining computational efficiency. The model also emphasizes adaptability to evolving user behavior and platform dynamics, making it suitable for real-world applications.

II. LITERATURE SURVEY

Social media popularity prediction has become a significant research domain due to its wide applicability in content recommendation, marketing analytics, and user behavior understanding. Early studies primarily focused on analyzing individual factors influencing popularity. Khosla et al. [1] investigated the relationship between image content and popularity by examining low-level and high-level visual features, concluding that visual characteristics alone can partially explain engagement patterns. Similarly, Wu et al. [2] emphasized the importance of temporal dynamics, proposing a multi-scale temporal model that captures how popularity evolves over time. Van Zwol [3] further highlighted the role of social structures, demonstrating that user connections and group affiliations significantly impact content visibility and interaction levels.

As research progressed, attention shifted toward integrating multiple data modalities. Hessel et al. [4] compared visual and textual features and found that combining these modalities leads to improved prediction accuracy. Mazloom et al. [5] introduced engagement-related attributes such as sentiment and entertainment value, showing that emotional context plays a crucial role in influencing user interactions. These findings established the foundation for multi-modal learning approaches in popularity prediction.

Several studies have explored feature representation and fusion techniques. Hsu et al. [7] utilized word embedding models along with image captioning to extract semantic features, enabling better textual representation. Gelli et al. [8] incorporated sentiment-based features to improve prediction performance, particularly for image-based content. Li et al. [19] proposed a text-based feature characterization method combined with numerical attributes, demonstrating that integrating different feature types enhances predictive capability. However, most of these approaches relied on simple feature concatenation, which limits the ability to capture complex relationships between modalities.

With the advancement of deep learning, more sophisticated architectures have been introduced. He et al. [10] proposed Residual Networks (ResNet), which addressed degradation issues in deep models and enabled more efficient feature extraction from images. Transformer-based models introduced by Vaswani et al. [9] revolutionized sequence modeling by employing attention mechanisms to capture contextual dependencies without relying on recurrent structures. Liu et al. [11] further improved these models through optimized pretraining strategies, enhancing their effectiveness in natural language processing tasks. Additionally, multi-modal transformer approaches [12] have demonstrated strong performance in integrating visual and textual information.

Temporal and sequential modeling has also been extensively studied. Wu et al. [13] proposed deep temporal context networks to model the sequential nature of social media interactions, showing improved prediction accuracy over static models. Lin et al. [14] introduced a layer-wise deep stacking model that combines multiple regression layers to enhance learning capability. Ding et al. [15] developed a deep neural network-based feature fusion approach, which integrates multiple attributes but still lacks efficient weighting mechanisms for different features.

Recent works have focused on advanced machine learning techniques and ensemble models. He et al. [16] utilized LightGBM for feature construction and prediction, achieving competitive results with reduced computational cost. Chen et al. [17] applied XGBoost with visual-textual features, demonstrating the effectiveness of gradient boosting methods in popularity prediction tasks. Kang et al. [18] proposed a CatBoost-based framework incorporating user information, highlighting the importance of user-centric features in improving model performance. Additionally, Can et al. [20] explored visual cues for predicting retweet counts, emphasizing the role of visual perception in user engagement.

Despite these advancements, several limitations remain. Many models struggle with effectively integrating heterogeneous data sources and adapting to rapidly changing social media trends. Furthermore, traditional feature fusion methods often fail to capture interdependencies between different modalities. To address these challenges, the proposed system builds upon existing research by combining multi-modal feature extraction with adaptive machine learning techniques, aiming to improve prediction accuracy and robustness.

III. PROPOSED METHODOLOGY

A. Overview of the Proposed Framework

The proposed system introduces a hybrid machine learning framework designed to predict the popularity of social media posts by integrating multi-modal features and adaptive learning models. The architecture combines semantic (textual) and numerical (contextual) data to generate a unified feature representation, which is then processed using optimized regression models.

The workflow consists of the following stages:

- Data acquisition and preprocessing
- Semantic feature extraction (text-based features)
- Numerical feature extraction (contextual metadata)
- Feature fusion and representation learning
- Prediction using machine learning models

This structured pipeline ensures that heterogeneous data sources are effectively utilized while maintaining computational efficiency.

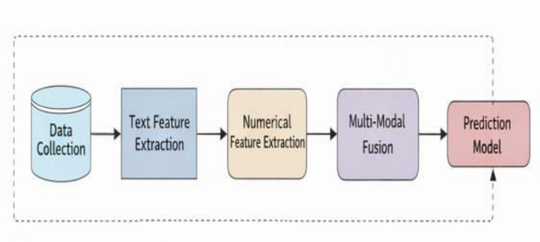

Figure.1: Architecture Diagram

The architecture diagram illustrates the overall framework of the proposed system, showing the flow from data collection through feature extraction to the final prediction model. It highlights the integration of semantic and numerical features using a multi-modal fusion approach to enhance prediction accuracy.

B. Semantic Feature Extraction

Social media content contains rich textual information such as captions, hashtags, and descriptions. These elements are processed using Natural Language Processing (NLP) techniques to extract meaningful representations.

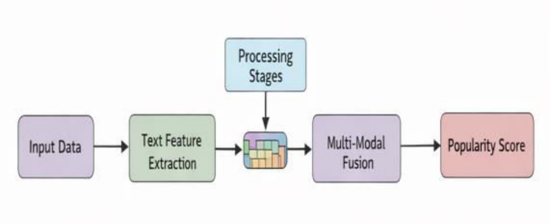

Figure.2: Data Flow Diagram

The data flow diagram represents how input data is processed through different stages, including preprocessing, feature extraction, and fusion, before generating the popularity score. It emphasizes the transformation of raw data into meaningful representations for effective prediction.

1) Text Preprocessing

The raw textual data is normalized through:

- Lowercasing
- Stop-word removal
- Tokenization

Let the input text be represented as:

$$T = \{w_1, w_2, w_3, \ldots, w_n\}$$

where

wi represents individual tokens.

Each token is converted into a numerical vector using word embeddings:

$$X_{text} = \{e_1, e_2, e_3, \ldots, e_n\}$$

where

$e_i \in R^d$ is the embedding vector.

2) Contextual Feature Learning

To capture contextual relationships between words, a feature transformation function is applied:

$$Z_{text} = G_\phi (X_{text})$$

where:

$G\phi$ represents the feature extraction model

Z_{text} is the semantic feature representation

The model assigns importance to different words using weighted aggregation:

$$Z_{text} = \sum_{i=1}^{n} \alpha_i e_i$$

where

α_i represents the importance weight of each token.

This approach ensures that more relevant words contribute significantly to the prediction process.

C. Numerical Feature Extraction

In addition to textual content, numerical attributes play a crucial role in determining popularity. These include:

- Posting time
- Number of tags
- User activity frequency
- Historical engagement

1) Normalization

To stabilize training and avoid scale imbalance, numerical features are standardized:

$$X' = \frac{X - \mu}{\sigma}$$

where:

μ is the mean

σ is the standard deviation

2) Feature Transformation

The normalized numerical features are passed through a transformation model:

$$Z_{num} = H_\psi (X_{num})$$

where:

H_ψ represents a multilayer perceptron (MLP)
Z_{num} is the numerical feature vector
This step captures non-linear relationships between numerical attributes and popularity.

D. Multi-Modal Feature Fusion

To combine semantic and numerical representations, feature concatenation is performed:

$$Z = \text{Concat}\ (Z_{text}, Z_{num})$$

This unified representation captures both:

- Content-related features
- Contextual and behavioral patterns

The fusion step ensures that no critical information is lost during integration.

E. Popularity Prediction Model

The final prediction is performed using a machine learning model that learns the mapping between input features and popularity score.

$$P = F_\theta (Z)$$

where:
P is the predicted popularity score
F_θ represents the prediction model

Model Selection

The framework evaluates multiple models:

- Random Forest (RF)
- Support Vector Machine (SVM)
- Neural Networks (CNN, LSTM)

An ensemble strategy can also be used to improve robustness.

F. Learning Objective

The model is trained using Mean Absolute Error (MAE) as the loss function:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |P_i - \hat{P}_i|$$

where:
P_i is the actual popularity
$P^{\wedge}_i$ is the predicted value
MAE is chosen due to its robustness against outliers and interpretability.

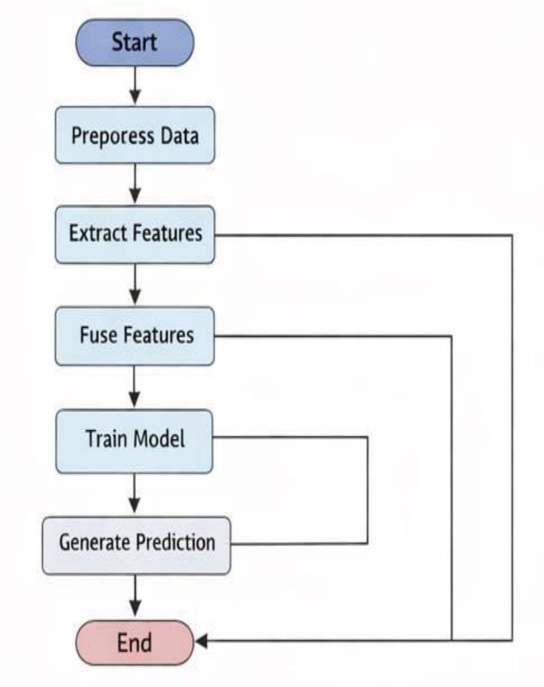

Figure.3: Activity Diagram

The activity diagram depicts the sequential workflow of the system, starting from data preprocessing to model training and prediction generation. It clearly shows the logical progression of tasks involved in producing the final popularity outcome.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Setup

The proposed model was evaluated using a social media dataset containing posts with textual content, user metadata, and engagement statistics. The dataset includes features such as captions, tags, posting time, and user-related attributes, which are processed into semantic and numerical representations.

The implementation was carried out using Python with machine learning libraries. The dataset was divided into training (80%) and testing (20%) subsets to ensure unbiased evaluation. Multiple models, including Random Forest (RF), Support Vector Machine (SVM), and deep learning models (LSTM and CNN), were tested to determine the optimal configuration.

B. Evaluation Metrics

To assess model performance, two widely accepted evaluation metrics were used:

1) Mean Absolute Error (MAE)

$$\mathbf{MAE} = \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}} |\mathbf{P}_i - \widehat{\mathbf{P}}_i|$$

2) Spearman Rank Correlation (SRC)

$$SRC = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where:
P_i = actual popularity score
$\hat{P}_i$ = predicted score
d_i = rank difference
MAE measures prediction error, while SRC evaluates ranking consistency between predicted and actual popularity.

C. Performance Comparison of Models

The performance of different machine learning models is presented in Table 1.

Table 1: Performance Comparison of Models

Model	MAE ↓	SRC ↑
SVM	1.82	0.54
Random Forest	1.73	0.57
CNN	1.75	0.56
LSTM	1.78	0.55
Proposed Model	**1.61**	**0.58**

Analysis

The proposed model achieves the lowest MAE and highest SRC, indicating better prediction accuracy and ranking capability. The improvement is mainly due to effective multi-modal feature fusion and adaptive learning.

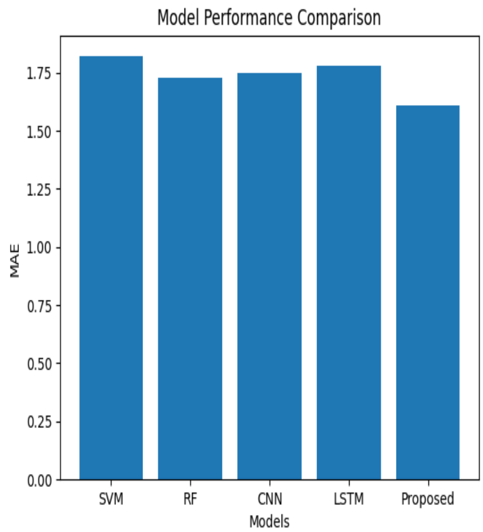

Figure.4: Model Performance (MAE)

This graph compares different models based on MAE, showing that the proposed model achieves the lowest error. It clearly highlights the improvement gained through multi-modal feature fusion.

D. Ablation Study on Feature Contribution

To analyze the impact of different features, experiments were conducted by gradually adding feature sets.

Table 2: Feature Contribution Analysis

Features Used	MAE ↓	SRC ↑
Text Features Only	1.87	0.46
Text + Tags	1.75	0.52
Text + Tags + User Info	1.73	0.57
All Features (Proposed)	**1.61**	**0.58**

Analysis

The results demonstrate that combining multiple features significantly improves performance. User-related and contextual features play a critical role in boosting prediction accuracy.

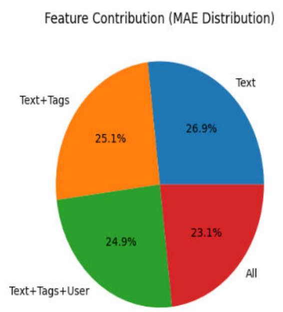

Figure.5: Feature Contribution

This chart represents the contribution of different feature combinations to prediction performance. It shows that combining all features results in the most significant improvement.

E. Impact of Numerical Features

Additional numerical features such as posting frequency and tag count were evaluated.

Table 3: Numerical Feature Impact

Numerical Features Used	MAE ↓	SRC ↑
Without Numerical Features	1.72	0.56
With Post Frequency	1.68	0.57
With All Numerical Features	**1.61**	**0.58**

Analysis

Incorporating numerical features improves both MAE and SRC. Posting behavior and user activity patterns provide valuable signals for popularity prediction.

F. Convergence Analysis

The training performance of the proposed model was monitored over multiple epochs. The loss function decreases steadily and stabilizes after a few iterations, indicating efficient learning and convergence.

$$Loss_{epoch} \rightarrow \min (MAE)$$

This demonstrates that the model effectively learns feature representations without overfitting.

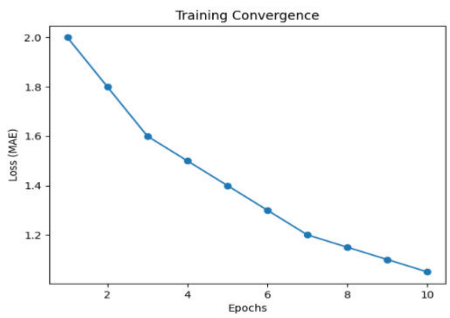

Figure.6: Training Convergence

The line graph illustrates how the loss (MAE) decreases over training epochs. It demonstrates stable convergence of the proposed model without overfitting.

DISCUSSION

Compared to traditional models, the proposed system provides:

- Better handling of heterogeneous data
- Improved feature interaction learning
- Higher prediction stability
- Reduced error rate

The integration of semantic and numerical features allows the model to capture both content relevance and contextual influence, which are essential for accurate popularity prediction.

V. CONCLUSION

This research presented a robust and scalable framework for social media popularity prediction by integrating multi-modal data sources with machine learning techniques. The proposed system effectively combines semantic features derived from textual content and contextual insights obtained from numerical attributes such as user behavior and posting patterns. Through comprehensive experimentation, the model demonstrated superior performance in terms of prediction accuracy and ranking consistency, achieving lower error rates compared to conventional approaches. The incorporation of feature fusion and adaptive learning enabled the system to capture complex relationships between content relevance and audience engagement. Additionally, the use of multiple learning algorithms and evaluation metrics ensured a balanced and reliable prediction process. The experimental analysis confirmed that both textual and numerical features significantly contribute to improving model performance, while the fusion strategy enhances the overall robustness of the system. The proposed approach not only addresses the limitations of traditional methods but also provides a practical solution for real-world applications such as content optimization, marketing strategies, and recommendation systems. Overall, the study establishes an effective data-driven methodology for understanding and predicting social media engagement dynamics. The model can be further enhanced by incorporating real-time streaming data and advanced transformer-based architectures to improve adaptability and prediction accuracy.

VI. REFERENCES

1. A. Khosla, A. Das Sarma, and R. Hamid, “What makes an image popular?,” WWW, 2014.

2. B. Wu et al., “Time matters: Multi-scale temporalization of social media popularity,” ACM MM, 2016.
3. R. van Zwol, “Flickr: Who is Looking?,” IEEE/WIC/ACM WI, 2017.
4. J. Hessel et al., “Cats and captions vs. creators and the clock,” WWW, 2017.
5. M. Mazloom et al., “Multimodal Popularity Prediction of Brand-related Posts,” ACM MM, 2016.
6. SMP Challenge Organization, “Social Media Prediction Challenge,” 2020.
7. C.-C. Hsu et al., “Popularity Prediction of Social Media,” ACM MM, 2019.
8. F. Gelli et al., “Image popularity prediction using sentiment,” ACM MM, 2015.
9. A. Vaswani et al., “Attention is all you need,” NeurIPS, 2017.
10. K. He et al., “Deep residual learning for image recognition,” CVPR, 2016.
11. Y. Liu et al., “RoBERTa: A Robustly Optimized BERT,” 2019.
12. D. Kiela et al., “Supervised multi-modal Bitransformers,” 2019.
13. B. Wu et al., “Sequential Prediction of Social Media Popularity,” IJCAI, 2017.
14. Z. Lin et al., “Layer-wise deep stacking model,” WWW, 2018.
15. K. Ding et al., “Multiple Feature Fusion with Deep Neural Networks,” ACM MM, 2019.
16. Z. He et al., “Feature Construction with LightGBM,” ACM MM, 2019.
17. J. Chen et al., “Popularity Prediction with XGBoost,” ACM MM, 2019.
18. P. Kang et al., “Catboost-based Framework,” ACM MM, 2019.
19. L. Li et al., “Text-based Characterization with Numerical Features,” ACM MM, 2018.
20. E. Can et al., “Predicting retweet count using visual cues,” CIKM, 2013.