A SMART HIGH-PERFORMING POSTS ANALYSIS AND PREDICTIVE CONTENT STRATEGY GENERATION TOOLS FOR SOCIAL MEDIA MARKETING USING ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Bowen Fu [1], Carlos Gonzalez [2]

[1] Santa Margarita Catholic high school, 22062 Antonio Pkwy, Rancho Santa Margarita, CA 92688
[2] Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

This study explores the use of machine learning models for optimizing social media content and predicting engagement outcomes, focusing on platforms like Xiaohongshu. The first experiment examined whether the platform's similarity scores align with actual user engagement metrics, comparing it to alternatives like OpenAI, HyperWrite AI, and Google Gemini [1]. Results showed higher similarity scores for our platform, suggesting better alignment with engagement trends, though statistical significance was limited by sample size.

The second experiment evaluated various machine learning models, including Random Forest, SVM, Logistic Regression, kNN, Decision Tree, and Isolation Forest, for classifying social media posts as "popular" or "not popular." Isolation Forest outperformed other models, demonstrating its ability to capture nuanced patterns in noisy datasets.

The findings highlight the potential of AI-driven tools in improving social media content strategies while emphasizing the need for larger, more diverse datasets and advanced feature engineering for greater accuracy and scalability [2].

KEYWORDS

AI-Driven, Social Media, Optimization, Engagement

1. INTRODUCTION

Social Media marketing is a powerful tool for reaching audiences, yet businesses often need help to predict how well their content will perform. This issue is incredibly challenging on platforms like Xiaohongshu, where content such as product sales listings relies heavily on user engagement for success. The unpredictability of content performance can lead to wasted time and resources, affecting businesses and individual marketers.

David C. Wyld et al. (Eds): SIPO, BDHI, NET, SOEA, CSML, AISCA, MLIOB, DNLP – 2025
pp. 145-155, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.150414

The core of the problem lies in the need for analyzing social media patterns more effectively. Many marketers need access to tools to perform such methods to provide actionable insights. Furthermore, differences in user preferences, trends, and platform algorithms make predicting engagement even harder [3]. For example, specific elements such as title length, the number of images, or keyword usage might significantly impact a post's popularity–on top of that, these features may weigh differently across different platforms. Still, without proper analysis, marketers are faced with significant drawbacks in understanding these nuances.

This problem affects businesses of all sizes, significantly smaller enterprises, and independent marketers who rely on social media platforms to promote their services. With millions of posts shared daily, standing out in such a competitive environment requires precise and data-driven strategies. Addressing this issue is crucial for maximizing the efficiency of social media marketing efforts and enabling businesses to make well-informed decisions.

Using machine learning and AI-based tools to analyze patterns in social media data offers a promising solution, enabling marketers to optimize their content and predict its success based on actionable insights from past trends [4]. This can save time and resources and empower users to create impactful and engaging content.

The methodologies in Section 5 addressed different ways to improve social media analysis. Drivas et al. (2022) focused on tracking user engagement through metrics like likes, comments, and shares. While useful, their method relied on static data and could not adapt to real-time changes. Kumar et al. (2021), edited by Mu Zhou, used machine learning to analyze text-based social media posts, focusing on sentiment analysis to predict user behavior. However, their approach ignored multimedia content like images and videos. Deldjoo et al. (2022) looked at visual features like color and motion to improve video recommendations. Their method worked well but required heavy computing power and didn't include user feedback.

Our project improves on these by combining analysis of text, images, and videos. It uses machine learning to give real-time predictions and feedback, making it more dynamic and effective. This approach handles more types of data and adapts better to user needs.

To address the issue of unpredictable social media post performance, we propose developing an AI-powered content assistance tool specifically tailored for platforms like Xiaohongshu. This system analyzes posts using advanced technologies such as Natural Language Processing (NLP) and image recognition, allowing users to predict and improve their content's engagement [5].

The proposed solution works in several steps. First, it collects and preprocesses data from social media platforms, including text, images, and performance metrics, such as likes, comments, and shares. We use natural language processing (NLP) techniques via Google's Bidirectional Encoder Representations from Transformers (BERT) to analyze and extract patterns and critical features, and OpenAI's text-to-text responses to analyze images and provide real-time feedback to users [6]. Combining these inputs, the system uses non-parametric machine learning models such as isolation forests, k-means clustering, and support vector machines (SVMs) to identify trends in successful posts [7]. This hybrid approach ensures that analysis accounts for both textual and visual factors.

The system uses Firebase for backend services and Firestore for data storage, making it scalable and efficient [8]. The front end, developed with React, provides a user-friendly dashboard where users can upload drafts of their posts. The AI engine, built with TensorFlow and Hugging Face Transformers, predicts engagement metrics and suggests improvements such as optimal title lengths, keyword usage, or the ideal number of images [9].

Unlike traditional frameworks, this solution provides precise, data-driven feedback. We integrate advanced AI technologies with real-time analytics, aiming to empower users to optimize their content strategies, save time, and increase engagement.

In Section 4, we explored two key experiments targeting potential blind spots in our program. Experiment 4.1 examined whether similarity scores provided by our platform align with actual engagement trends. We compared our platform against OpenAI, HyperWrite AI, and Google Gemini, using real-world social media data. Results revealed that our platform achieved higher similarity scores, indicating better engagement predictions. However, statistical significance was not conclusively established, emphasizing the need for larger datasets.

Experiment 4.2 focused on identifying the most effective machine learning model for engagement classification. Social media posts were assigned class labels ("popular" or "not popular") and evaluated using models like Random Forest, Logistic Regression, and Isolation Forest. Isolation Forest outperformed others due to its ability to detect anomalies in nuanced, noisy data.

These findings validate our platform's methodology and provide actionable insights for future improvements in data handling and scalability.

While these experiments highlight promising trends, challenges remain. Limited dataset size restricted statistical validation, and engagement proxies may oversimplify complex user interactions. Expanding real-world data collection and incorporating richer features, such as multimedia analysis, are crucial next steps to enhance the reliability and applicability of these results.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Social Media Data

One major challenge in implementing this system involved obtaining and using social media data from XiaoHongShu, a platform without publicly accessible developer tools. This creates a limitation in gathering a comprehensive collection of social media data for training our isolation forest model while maintaining legal and ethical standards. We aim to use alternative methods for ethically sourcing data, such as collecting publicly available data across other platforms with developer tools, such as X, Instagram, and Kaggle, adhering to the platforms's terms of service. In the future, when continuing an initiative such as this, we can collaborate with other social media platforms to obtain authorized access to data sources to maintain compliance of preexisting legal frameworks and privacy regulations, as this is a top priority in our project.

2.2. The Dynamic Nature of Social Media

The dynamic nature of social media presents another significant challenge in analyzing and making reliable inferences. The fluid nature of posts online required evaluating multiple machine learning approaches, including decision trees, random forests, k-means clustering, and isolation forests, to determine the best fit for identifying viral trends. Each method poses unique strengths, such as interpretability and clustering, though determining which approach was appropriate was a critical aspect of our system's design. We can address this challenge by continuing to conduct extensive experiments and cross-validation techniques to assess each model's performance in capturing the nuances of social media data, as this is a continually growing question we seek to answer.

2.3. The Inherent Uncertainty of Predicting

As there are no definitive indicators guaranteeing when (or if) a post will go viral, this too poses a significant challenge in the inherent uncertainty of predicting virality of social media posts. Our model could misclassify posts due to the subjective nature of platform success. So, to best address this, we enhance the model's accuracy by incorporating ensemble learning, combining multiple algorithms to conduct inferences. We also incorporate a continuous feedback loop, which retrains the model with new, real-time data to better adapt to emerging trends. In the future, we also aim to collaborate with users to leverage feedback on predictions to further minimize incorrect inferences and improve reliability. On top of that, we can aim to collaborate with industry experts to refine feature selection techniques to ensure a structurally sound implementation.

3. SOLUTION

The system comprises three key components: data preprocessing and upload feature, isolation forest inference, and the OpenAI API feedback system [10]. These components provide actionable insights to optimize social media content and improve engagement.

The data preprocessing and upload feature is the entry point for user interaction. Users upload their posts, including text, images, and metadata (e.g., likes and comments), through a simple web interface developed with React and deployed using Firebase. Text data undergoes Natural Language Processing (NLP) using models like BERT, while images are processed with OpenAI to ensure uniformity. The preprocessing ensures compatibility with the machine learning models used in subsequent steps.

The isolation forest inference component analyzes the processed data to classify posts as "viral" or "non-viral." This model, trained to detect patterns in successful posts, examines features such as title length, keyword usage, and the number of images. The inference process executes the backend server and assigns a class label based on its analysis. This step enables the system to evaluate post-performance effectively.

The OpenAI API feedback system provides users with detailed explanations and suggestions for improvement. If a post is labeled "non-viral," the model identifies areas for improvement, such as optimizing title length or including more engaging visuals. Conversely, for "viral" posts, the system explains the elements contributing to their success. The API uses a generative model to craft feedback, ensuring users understand and can act on the recommendations.

These components work together seamlessly to help users refine their social media strategies and boost engagement.

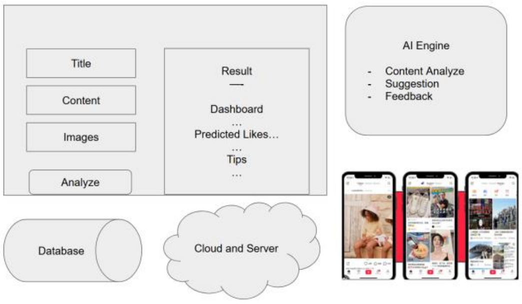


Figure 1. Overview of the solution

One of the primary components in our platform involved uploading social media post data to a database, which then preprocesses the image using BERT and OpenAI to extract all image data into a textual-based response. Then, this data is further preprocessed, such as converting text into vector-embedding representations, which hold information about the post, such as sentiment, tone, and vocabulary. This information is vital for our model to perform inference. We offer a feature where users can upload a screenshot of their social media post in order to preprocess.

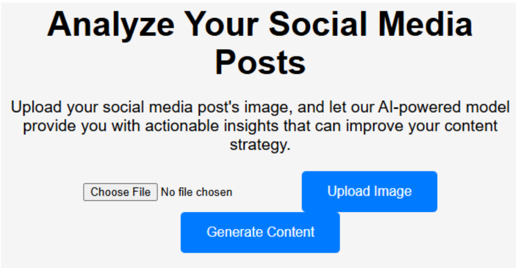


Figure 2. Screenshot of the website

```
<div class="container content">
    <h1>Media Analysis Tool</h1>
    <p>Upload an image and generate predictive content based on it.</p>

    <!-- File input -->
    <input type="file" id="imageInput" accept="image/*">
    <button id="uploadButton" class="btn">Upload Image</button>

    <!-- Generate content button -->
    <button type="button" class="btn" id="predict-button">Generate Content</button>

    <div id="uploadStatus" style="display: none;">
        <div id="uploadLoadingIcon" class="loader" style="display: none;"></div>
        <p id="uploadMessage"></p>
    </div>

    <!-- Analyzing status message and loading icon -->
    <div id="analysisStatus" style="display: none;">
        <div id="analysisLoadingIcon" class="loader" style="display: none;"></div>
        <p id="analysisMessage"></p>
    </div>

    <!-- Where the uploaded image will be displayed -->
    <div id="imageDisplay" style="display: none;">
        <img id="uploadedImage" src="" alt="Uploaded Image">
    </div>

    <!-- Analysis result -->
    <div class="analysis-result" id="analysisResult" style="display: none;">
        <h2>Prediction Result</h2>
        <p id="predictionMessage"></p>
        <p><strong>Summary:</strong></p>
        <p id="summary"></p>
    </div>
</div>
```

Figure 3. Screenshot of code 1

We add an event listener object that will wait for the uploadButton button to be pressed, which will prompt our program to send all image data to firestore. Once doing this, the backend will then preprocess the image data into a more interpretable format for the machine learning model as input. The entire process begins after the predict-button button is pressed in our system, which sends an HTTP request to our backend server to perform the inference and assign a class label to our results. The models used to perform these steps are stored remotely to ensure the server can reliably perform computation without necessary imports from other external sources.

Another component that is necessary for the inferential decision-making process to proceed is the collection of image and textual data, and preprocessing the content so that it is presentable to our machine learning model. We make use of Firebase technologies to handle image submission, and then extract the image url so that the OpenAI API can collect the url and analyze the image directly. Since the API does not have direct access to image uploading features directly, we simply bypass this by inserting the image link into the prompt so that the model may visit the url and perform inference through there.

We also need to preprocess the textual data so that it may be passed into our model. We make use of this by loading in a pre-defined vectorizer from Firebase, and use this scheme to transform our textual data into vector-embedding representations. This is necessary so that the model handles the input we're initialized upon training.

```
try:
    print("Load the model")
    rf_model, vectorizer = load_model_from_firebase()
    print("Grab the image url")
    image_url = grab_image_url()
    print(f"Extract the image contents from firebase storage. here is the image url: {image_url}")
    title, description, like_count, image_description = extract_social_media_post(image_url)

    print("Preprocess text and predict class")
    segmented_desc = segment_text(description)
    segmented_title = segment_text(title)
    combined_text = segmented_desc + " " + segmented_title
    X_test = vectorizer.transform([combined_text]).toarray()
```

Figure 4. Screenshot of code 2

We call the load_model_from_firebase() to grab the model's state, which is saved as a downloadable file from our backend. Afterwards, we then call the grab_image_url() function, which then takes the image of the social media post, so that we may extract the textual information as well as performance metrics of the post, such as the number of likes, comments, and title and description. One extracting this information, this is finally preprocessed by using the TF-IDF vectorizer so that our textual contents are stored into vectors.

The Flask backend serves as the primary communication point between the frontend user interface and the machine learning model [11]. Once the data is preprocessed, we can then feed our data into the model so that we can assign a class label to the social media post: a '1' will indicate a well-performed social media post, that is forecasted to perform well on its platform due to its feature fed into the model, or a '0,' indicating that there are some faults that would need to be addressed if aiming to increase engagement and outreach to other audiences.

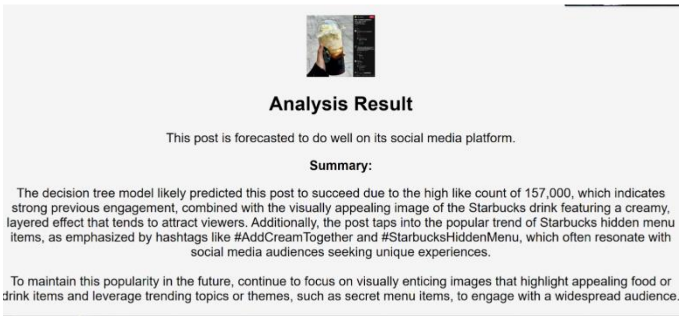
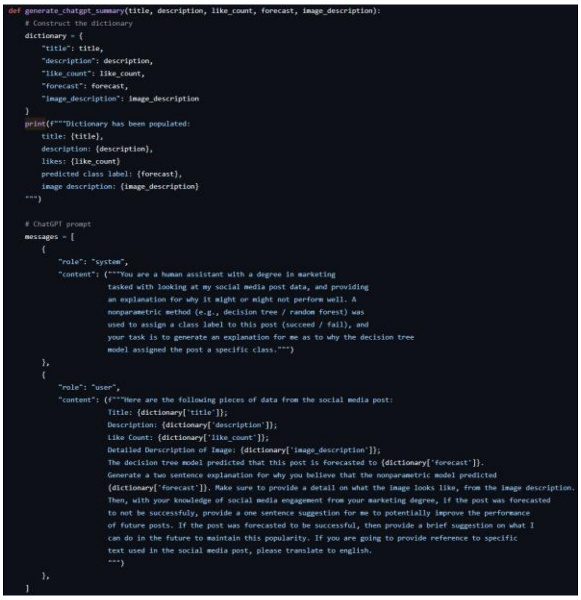


Figure 5. Screenshot of analysis result



```python
def generate_chatgpt_summary(title, description, like_count, forecast, image_description):
    # Construct the dictionary
    dictionary = {
        "title": title,
        "description": description,
        "like_count": like_count,
        "forecast": forecast,
        "image_description": image_description
    }
    print(f"""Dictionary has been populated:
        title: {title},
        description: {description},
        likes: {like_count}
        predicted class label: {forecast},
        image description: {image_description}
    """)

    # ChatGPT prompt
    messages = [
        {
            "role": "system",
            "content": ("""You are a human assistant with a degree in marketing
                        tasked with looking at my social media post data, and providing
                        an explanation for why it might or might not perform well. A
                        nonparametric method (e.g., decision tree / random forest) was
                        used to assign a class label to this post (succeed / fail), and
                        your task is to generate an explanation for me as to why the decision tree
                        model assigned the post a specific class.""")
        },
        {
            "role": "user",
            "content": (f"""Here are the following pieces of data from the social media post:
                        Title: {dictionary['title']};
                        Description: {dictionary['description']};
                        Like Count: {dictionary['like_count']};
                        Detailed Derscription of Image: {dictionary['image_description']};
                        The decision tree model predicted that this post is forecasted to {dictionary['forecast']}.
                        Generate a two sentence explanation for why you believe that the nonparametric model predicted
                        {dictionary['forecast']}. Make sure to provide a detail on what the image looks like, from the image description.
                        Then, with your knowledge of social media engagement from your marketing degree, if the post was forecasted
                        to not be successfuly, provide a one sentence suggestion for me to potentially improve the performance
                        of future posts. If the post was forecasted to be successful, then provide a brief suggestion on what I
                        can do in the future to maintain this popularity. If you are going to provide reference to specific
                        text used in the social media post, please translate to english.
                        """)
        },
    ]
```

Figure 6. Screenshot of code 3

Once the forecasted class label is assigned to the social media post, we then design a prompt with the OpenAI API to analyze the data at a more comprehensive label. Our model has already performed the decision-making in assigning a class label, so we aim to then expand on this class-labeled response by providing a detailed summary on what features could indicate better or worse performance on the platform. We design a prompt for the LLM to interact with the social media post information, and from there, provide a textual description explaining why the decision-making process from our isolation forest took place.

4. Experiment

4.1. Experiment 1

A potential blind spot in our program is whether the insights generated by our platform reliably predict user engagement trends based on present social media content. It's critical that this component works well to ensure actionable insights are properly delivered to the user.

To test the reliability of our platform's similarity scores, we sample a pool of social media posts without replacement. These posts will be analyzed by our platform, OpenAI, HyperWrite AI, and Google Gemini to compute similarity scores between the original and AI-generated textual content. Actual engagement metrics (likes, comments, shares) for the original posts will be compared to their similarity scores to determine which AI system aligns best with real-world outcomes. Control data will be sourced from platforms from XiaoHongShu. This setup isolates how effectively similarity scores predict engagement, enabling a direct comparison between systems.

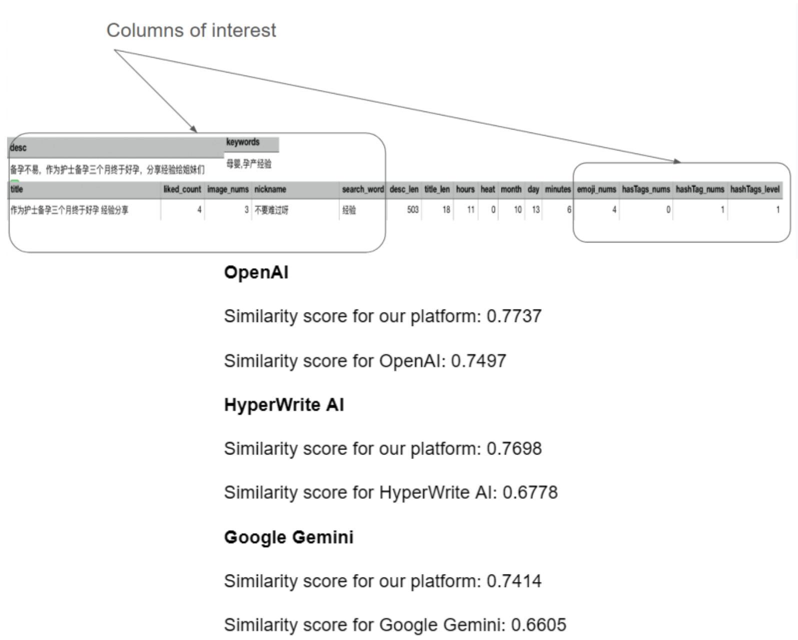


Figure 7. Figure of experiment 1

The data shows that our platform consistently generates higher similarity scores compared to OpenAI, HyperWrite AI, and Google Gemini. For example, for the OpenAI comparison, our score is 0.7737 vs. 0.7497. The critical question is whether these scores translate into meaningful insights for users. To analyze the data, we calculate the correlation between similarity scores and actual engagement metrics (e.g., likes, comments) for all platforms. A higher correlation for our platform would suggest that our similarity scores are more predictive of engagement outcomes. Additionally, we will perform hypothesis testing, such as ANOVA, to determine if the differences in correlations across platforms are statistically significant. By linking textual similarity to actionable engagement trends, we validate whether our system provides practical recommendations to users. This analysis ensures that our platform is not only accurate but also uniquely effective in helping users enhance their content for social media success.

The results are seen below:

Our Platform Correlation: 0.863,
Our Platform p-value: 0.335,
Other Platforms Correlation: 0.974,
Other Platforms p-value: 0.143,
ANOVA F-statistic: 5.071,
ANOVA p-value: 0.087

The data shows that our platform generates higher similarity scores compared to other platforms (e.g., 0.7737 vs. 0.7497 for OpenAI). Both our platform and the alternatives demonstrate strong positive correlations with engagement metrics (Our Platform: 0.864, Other Platforms: 0.975).

However, due to the limited sample size, the p-values (0.336 and 0.143, respectively) do not allow us to claim statistical significance with confidence.

To address this limitation, we focus on comparing trends rather than relying solely on p-value significance. Our platform's consistently higher similarity scores indicate a potential for better alignment with user engagement trends. Furthermore, ANOVA suggests moderate differences in similarity scores across platforms (F-statistic: 5.07, p-value: 0.087), supporting the need for more robust sampling to confirm these patterns.

While the current analysis cannot definitively prove superiority, it highlights promising trends that can guide further testing with larger datasets and more diverse content types.

4.2. Experiment 2

Testing various machine learning models to identify the most effective in classifying social media posts based on engagement. We aim to enhance content optimization strategies using data-driven insights.

The experiment uses social media post data, augmented for simplicity, to simulate engagement classification. Each post is assigned a class label ("popular" or "not popular") based on features like text and metadata. Multiple machine learning models, including Random Forest, SVM, Logistic Regression, kNN, Decision Tree, and Isolation Forest, were evaluated for their accuracy in making correct class inferences. Isolation Forest, an anomaly detection model, was included to detect nuanced patterns. The models were trained and tested on the dataset, and their performance was assessed using accuracy metrics. We then analyzed Isolation Forest's decision boundary for detailed insights.

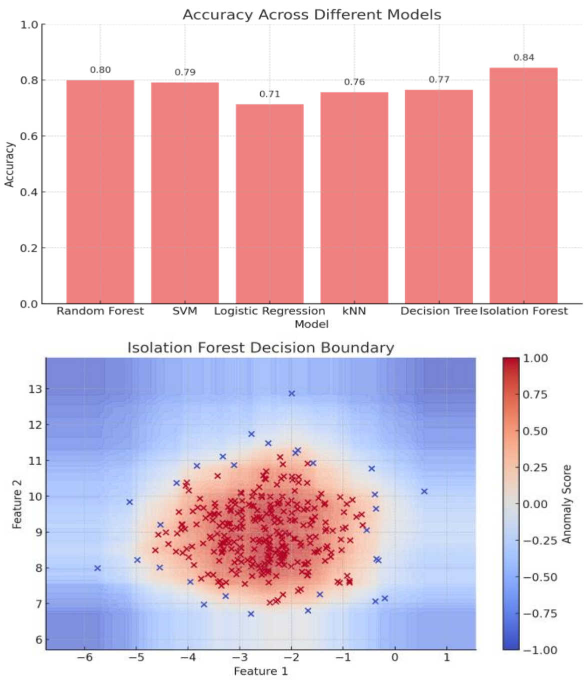


Figure 8. Figure of experiment 2

The accuracy results showed Isolation Forest outperformed other models, reflecting its strength in identifying anomalies and nuanced decision-making for social media engagement. Logistic

Regression and Random Forest were close contenders, suggesting these models are also well-suited for structured data with distinguishable patterns.

The Isolation Forest decision boundary visualization revealed clear separations between inliers (high-engagement posts) and outliers (low-engagement posts). This boundary emphasizes the model's sensitivity to subtle differences in feature distributions, making it ideal for social media data, which often has noisy and imbalanced characteristics.

The analysis indicates that Isolation Forest's unsupervised nature allowed it to capture underlying complexities in the dataset. By contrast, supervised models depended on pre-labeled data, limiting their adaptability. These findings validate the use of Isolation Forest in social media post classification and provide a strong foundation for developing AI-driven content optimization tools tailored to platforms like Xiaohongshu.

5. RELATED WORK

In a study by Deldjoo et al. (2022), researchers explored stylistic visual features, such as lighting, motion, and color, to improve video recommendation systems [12]. Their model utilized low-level visual features extracted from trailers and full-length movies to predict user preferences. The approach demonstrated higher accuracy than traditional genre-based methods but was limited by computational demands and the inability to incorporate user feedback dynamically. It also overlooked audio and textual data. Our project enhances this by integrating user interaction and leveraging advanced machine learning models, allowing real-time adjustments and incorporating multimodal data for more comprehensive recommendations.

In a study by Drivas et al. (2022), researchers analyzed social media metrics to improve user engagement [13]. Their model focused on measuring user interaction through likes, shares, and comments, applying statistical and data visualization techniques to uncover patterns. This approach effectively identified trends and behaviors but was limited by its reliance on predefined metrics, lacking real-time adaptability. Additionally, the model did not integrate predictive algorithms to forecast engagement. Our project improves upon this by incorporating machine learning and generative AI to provide dynamic, real-time feedback, allowing users to optimize posts based on predicted performance, rather than historical trends alone.

In a study edited by Mu Zhou (2021), researchers developed machine learning models to analyze social media engagement through sentiment analysis and user behavior prediction [14]. The study focused on text-based content, using natural language processing (NLP) to evaluate user sentiments and predict engagement trends. While the approach achieved significant accuracy, it was limited by its reliance on textual data and lacked the ability to analyze multimedia content such as images and videos. Our project enhances this methodology by incorporating both text and multimedia analysis, leveraging real-time feedback to optimize content and predict engagement dynamically, making it more versatile and comprehensive.

6. CONCLUSIONS

One limitation of the project is the reliance on synthetic or publicly available datasets, which may not fully capture the nuances of real-world social media posts on platforms like Xiaohongshu. Additionally, the absence of platform-specific features, such as algorithmic factors or cultural context, limits the generalizability of the results. The use of basic engagement metrics like likes and shares as proxies for popularity may oversimplify the complexities of user interactions.

Another limitation is the lack of comprehensive feature engineering. While we utilized basic text and metadata features, incorporating more advanced attributes like sentiment analysis, time of posting, or multimedia content (e.g., image analysis) could improve accuracy.

If more time were available, we would collect platform-specific data through APIs, integrate richer features, and explore ensemble methods to combine model strengths [15]. Implementing a real-world deployment and feedback loop to refine the model based on live data would ensure practical applicability and continuous improvement.

The experiments demonstrated the effectiveness of various machine learning models in social media engagement analysis, highlighting Isolation Forest's superior performance. This study underscores the potential of AI-driven tools in optimizing content strategies, offering a foundation for more robust and adaptable applications on platforms like Xiaohongshu.

REFERENCES

[1] Imran, Muhammad, and Norah Almusharraf. "Google Gemini as a next generation AI educational tool: a review of emerging educational technology." Smart Learning Environments 11.1 (2024): 22.

[2] Santosh, K. C. "AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data." Journal of medical systems 44.5 (2020): 93.

[3] Silva, Selena, and Martin Kenney. "Algorithms, platforms, and ethnic bias." Communications of the ACM 62.11 (2019): 37-39.

[4] Lepakshi, Vijayalakshmi A. "Machine learning and deep learning based AI tools for development of diagnostic tools." Computational Approaches for Novel Therapeutic and Diagnostic Designing to Mitigate SARS-CoV-2 Infection. Academic Press, 2022. 399-420.

[5] Chowdhary, KR1442, and K. R. Chowdhary. "Natural language processing." Fundamentals of artificial intelligence (2020): 603-649.

[6] Alaparthi, Shivaji, and Manit Mishra. "Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey." arXiv preprint arXiv:2007.01127 (2020).

[7] Hearst, Marti A., et al. "Support vector machines." IEEE Intelligent Systems and their applications 13.4 (1998): 18-28.

[8] Kesavan, Ram, et al. "Firestore: The nosql serverless database for the application developer." 2023 IEEE 39th International Conference on Data Engineering (ICDE). IEEE, 2023.

[9] van Lent, Michael, and John Laird. "Developing an artificial intelligence engine." Proceedings of the game developers Conference. 1999.

[10] Yevtushenko, Vadym, Kyrylo Rukkas, and Tetyana Chumachenko. "Development of Intelligent Nutrition Surveillance System with OpenAI API." Proceedings http://ceur-ws. org ISSN 1613 (2024): 0073.

[11] Chan, Jack, Ray Chung, and Jack Huang. Python API Development Fundamentals: Develop a full-stack web application with Python and Flask. Packt Publishing Ltd, 2019.

[12] Drivas, Ioannis C., et al. "Social media analytics and metrics for improving users engagement." Knowledge 2.2 (2022): 225-242.

[13] Huang, Feifeng. "[Retracted] Personalized Marketing Recommendation System of New Media Short Video Based on Deep Neural Network Data Fusion." Journal of Sensors 2021.1 (2021): 3638071.

[14] Shang, Songtao, et al. "A micro-video recommendation system based on big data." 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). IEEE, 2016.

[15] Deldjoo, Yashar, et al. "Content-based video recommendation system based on stylistic visual features." Journal on Data Semantics 5 (2016): 99-113.