International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

WEB PAGE USER INTERACTION ON SOCIAL MEDIA

2111CS020348-K Praneeth 2111CS020349-P Pranay

2111CS020350-T Pranay 2111CS020351-V Pranay

2111CS020353-G Praneeth Guide: Prof T Ramya

Abstract: The study explores the usefulness of incorporating both image and text data to predict user engagement on social media. The research collected *350k Facebook posts* and incorporated five types of machine learning models: text-based Neural Networks (NN), image-based Convolutional Neural Networks (CNN), Word2Vec, decision trees, and a combination of text-based NN and imagebased CNN. The models are unique in their use of the data. The models learn and test on advertisement posts in order to predict user engagement, which includes share count, comment count, and comment sentiment. The study found that *combining image and text data produced the best models*. The research further demonstrates that combined models outperform random models.

1 INTRODUCTION

Gathering data included image processing. Part of image processing is denoising images in order to emphasize important image features. The features deemed important vary for each denoising algorithm. Generally, the denoising algorithms emphasize edges, gradient contrasts, and curves. One result of denoising can be a decreased in the amount of image data, which saves computer space. The reduction in image size makes storing all the images on a machine more practical. While reducing data size is helpful, the ultimate goal is not reducing image size. The ultimate goal is eliminating image noise that might interfere with image-based model training. By eliminating noise, the models can focus on image aspects that denoising has emphasized Part of denoising is preparing the image for analysis. Many of the machine learning models require for all images to be the same size. Therefore, reducing image size is important for the final machine learning model.

1.1 PROBLEM STATEMENT

The thesis focuses on predicting user engagement in social media, particularly on Facebook posts with image and text content. It addresses the challenges of user dissatisfaction with ads affecting engagement and revenue. Objectives include developing predictive models, offering insights to companies, and fostering advertiser confidence. The research analyzes 350k Facebook posts, utilizing machine learning to predict engagement metrics. Ultimately, the thesis aims to advance understanding and provide practical solutions for optimizing advertising effectiveness on social media platforms.

1.2 TECHNIQUES USED

Image Denoising: The research incorporated image processing techniques including denoising algorithms to enhance important image features and reduce noise. This involved emphasizing edges, gradient contrasts, and curves in images to prepare them for analysis.

Data Preprocessing: The dataset was preprocessed to remove irrelevant features and noise from images, ensuring that the data used for training the machine learning models was clean and focused on relevant aspects.

Standardization of Image Size: To ensure compatibility with machine learning models, all images were standardized to the same size. This

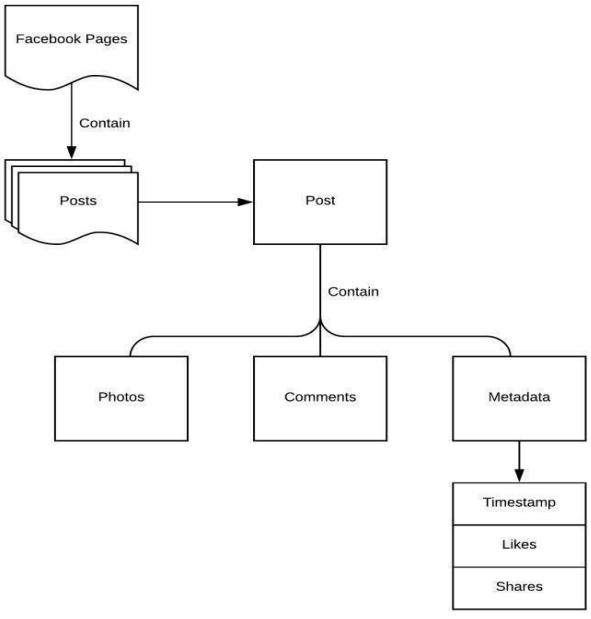
standardization is crucial for consistency in training and testing image-based models.

Machine Learning Models: Five types of machine learning models were utilized: Text-based Neural Networks (NN), Image-based Convolutional Neural Networks (CNN), Word2Vec, Decision Trees, Combination of text-based NN and image-based CNN

Model Evaluation Metrics: Various evaluation metrics were used to assess model performance, including: Mean Squared Error (MSE), Mean Absolute Error (MAE), Accuracy metrics for binary predictions.

Combined Text and Image Data: The study emphasized the usefulness of combining text and image data for predicting user engagement on social media. This approach was found to produce the best models compared to using text or image data alone.

1.3 ARCHITECTURE



1.4 DATASET DESCRIPTION

The dataset used in this study comprises 800 advertisements from Facebook, Twitter, and LinkedIn. It includes ad content such as text and images, as well as metadata like the number of likes, shares, and comments. The data was preprocessed to remove irrelevant features and denoise the image data. Ultimately, 600 ads were used for training the machine learning models, and 200 ads were used for testing. This dataset served as the foundation for predicting user interaction on social media using machine learning techniques.

1.5 MODEL EVALUATION AND METRICS

The model evaluation metrics used in this study included regression metrics for continuous prediction, such as mean squared error (MSE) and mean absolute error (MAE). Additionally, the study employed binary prediction to compare the performance of different advertisements, and the accuracy of these binary predictions was used as a measure of model success. The research also compared the created models with a random model to demonstrate if the models outperformed random guessing. These evaluation metrics were essential in assessing the performance of the machine learning models in predicting user interaction on social media.

2 LITERATURE REVIEW

Text-Based Models: Review existing literature on text-based machine learning models used to predict user engagement on social media. Discuss the strengths and limitations of text-based models in capturing user interaction.

Image-Based Models: Explore previous research on image-based machine learning models for predicting user engagement.

Highlight the role of image features and their impact on user interaction.

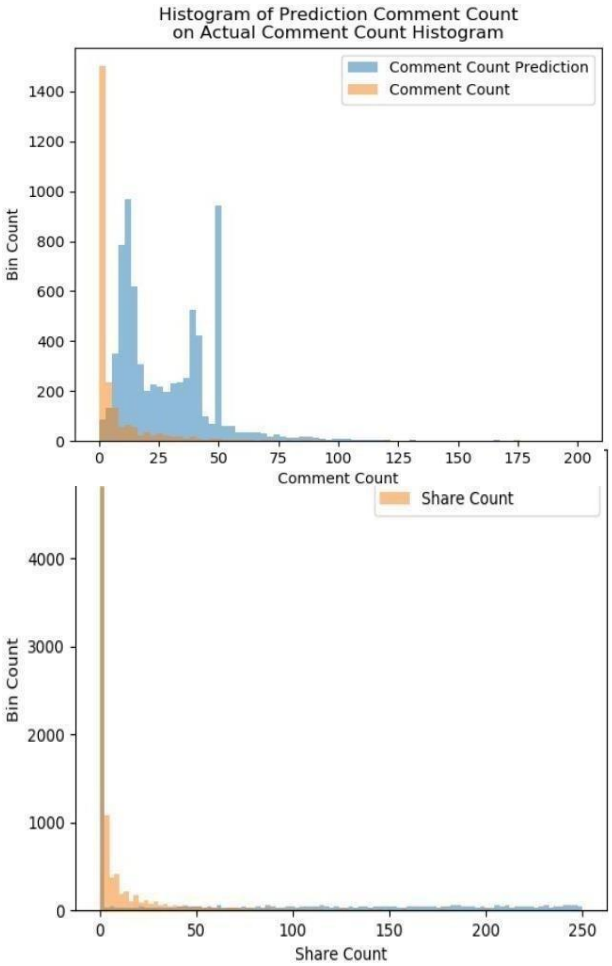
Combined Text and Image Models: Examine studies that have integrated both text and image data to

predict user engagement. Assess the effectiveness of combined models compared to text-only or image-only approaches.

Denoising Algorithms: Discuss the use of denoising algorithms in preprocessing images for user engagement prediction. Evaluate different denoising techniques and their effects on model performance.

Standardization of Image Data: Review literature on standardizing image size and format for machine learning model compatibility. Highlight the importance of image data consistency in predictive modeling.

3 EXPERIMENTAL RESULTS



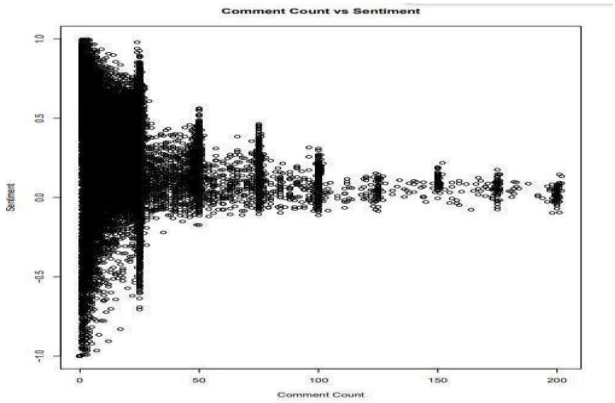


FIGURE 6.5: Plot of Comment Count vs Comment Sentiment

4 CONCLUSION

This research extends existing knowledge on user interaction prediction by investigating diverse machine learning models. It

explores the potential of combined image-based and text-based models, highlighting the superiority of such fusion models in predicting user engagement. The study establishes that these combined models outperform individual text-based and image-based models across various metrics. Moreover, the research applied these combined models to real-world scenarios, specifically in predicting user engagement for advertisement posts. It successfully generated predictions for comment count, share count, and comment sentiment, which were then compared with actual engagement metrics. The models exhibited accuracy rates of 57% for comment sentiment, 55% for comment count, and 53% for share count, surpassing the performance of random models (50%). These outcomes provide a robust foundation for future studies aiming to enhance user interaction prediction methodologies, particularly when leveraging combined models in practical applications and real-world scenarios. The study serves as a benchmark, guiding further exploration and development in this domain.

5 FUTURE WORK

Certainly, here is a revised version focusing on the future scope of the project: The research highlights the superior performance of image-based machine learning models compared to text-based models, particularly in predicting user engagement on social media platforms like Facebook. Image models demonstrated significant advantages, surpassing text models by over 40% in comment sentiment analysis and 3.5 times in estimating share counts. Future prospects involve leveraging Word2Vec for sentiment prediction, calling for a broader dataset to refine the models effectively. The study aims to tackle challenges in gathering data from the Facebook API to enhance the overall model performance. Further project development includes integrating diverse models to refine predictions for user interaction. There's a suggestion of potential relationships between metrics such as share count and comment sentiment, which might offer predictive insights between them. Moreover, successful advertisement using specific metrics could potentially indicate success in other areas. hese models extend beyond prediction, enabling content recommendations and suggesting transformations to enhance image performance on social media. The objective is to improve image engagement using filters, transformations, and rotations, aiming for better traction on social media platforms.

6 REFERNCES

Ahmadi, R. et al. (2013) 'Survey of image denoising techniques', Life Science Journal, 10(1), pp. 753–755. doi: 10.1.1.107.1823.

Aittala, M., Aila, T. and Lehtinen, J. (2016) 'Reflectance modeling by neural texture synthesis', ACM Transactions on Graphics, 35(4), pp. 1–13. doi: 10.1145/2897824.2925917.

Buduma, N. (2015) Fundamentals of Deep Learning, CEUR Workshop Proceedings. doi: 10.1017/CBO9781107415324.004.

Casari, A. (2017) 'Mastering Feature Engineering Principles and Techniques for Data Scientists'. Oreilly Associates Inc.

Chen, T. et al. (2014) 'DeepSentiBank: Visual Sentiment Concept Classification with Deep Convolutional Neural Networks'. Available at: http://arxiv.org/abs/1410.8586. "Click-Through Rate Prediction." Kaggle Competition (2015), https://www.kaggle. com/c/avazu-ctr-prediction.

Chu, T. C. et al. (1985) 'Applications of Digital . Image-Correlation Techniques to Experimental Mechanics', (August 2015). doi: 10.1007/BF02325092.