Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250402025
Volume 4, Issue2 (May-August 2025), PP: 199-205.
www.indjcst.com
ISSN No: 2583-5300

Learning Mechanisms without Experimenting: Examining Using Dataset

Saravanan S[1], Komal Dabber[2], Priyanka S Metgud[3], Hita Lakshmi[4]

[1]Assistant Professor, Department of Computer Science Engineering in Data Science, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India.

[2,3,4] Students, Department of Computer Science Engineering in Data Science, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India.

To Cite this Article: *Saravanan S[1], Komal Dabber[2], Priyanka S Metgud[3], Hita Lakshmi[4], "Learning Mechanisms without Experimenting: Examining Using Dataset", Indian Journal of Computer Science and Technology, Volume 04, Issue 02 (May-August 2025), PP: 199-205.*

Abstract: *This research explores a data-driven alternative to traditional machine learning experimentation by utilizing publicly available datasets instead of conducting resource-intensive and time-consuming physical studies. The objective is to investigate the predictive performance and interpretability of multiple regression algorithms, namely Linear Regression, Ridge Regression, Lasso Regression, and Decision Tree Regression, on real-world data. These models are evaluated based on their ability to uncover meaningful patterns, relationships, and potential causal inferences within the dataset. Emphasis is placed on essential preprocessing steps, including data cleaning, transformation, and encoding, to ensure the quality and consistency required for reliable model training and evaluation. By comparing these algorithms across common performance metrics such as mean squared error, mean absolute error, and R^2 score, the study provides insights into their robustness and generalizability. This approach facilitates repeatable and scalable experimentation, supporting rapid hypothesis testing and model optimization without the logistical constraints of physical data collection. Overall, the methodology contributes to accelerating machine learning research by promoting efficient, cost-effective practices while maintaining scientific rigor.*

Key Words: *Machine learning, dataset analysis, social media analytics, hypothesis testing, observational analysis.*

I.INTRODUCTION

Machine learning (ML) has significantly transformed the way systems analyze, learn, and make predictions based on data. These advancements have enabled automation in various domains, including healthcare, finance, and social media. However, the reliability and efficiency of ML models heavily depend on the datasets used for their development. Traditionally, experimental methodologies have been the cornerstone of validating ML models, involving rigorous training, testing, and benchmarking processes. These experiments help assess model accuracy, generalizability, and robustness through statistical metrics such as accuracy, precision, recall, F1-score, and ROC curves.

While empirical validation remains a crucial aspect of ML research, it comes with several challenges. Conducting experiments can be resource-intensive, time-consuming, and computationally expensive. Many experimental setups require vast amounts of labeled data and high-performance computing resources, which are not always readily available. Moreover, certain ML applications, such as those involving medical data, climate predictions, or large-scale financial modeling, face ethical and logistical constraints, making direct experimentation impractical.

Another significant challenge in traditional experimentation is data-related biases. Issues like class imbalance, noise, and non-representative datasets can lead to misleading conclusions and poor generalization in real-world applications. For instance, an ML model trained on a dataset with skewed class distributions may develop a bias toward the dominant class, reducing its effectiveness when applied to diverse data. Methods like SMOTE (Synthetic Minority Over-Sampling Technique), Tomek Links, and Wilson's Edited Nearest Neighbor Rule have been developed to mitigate these issues, yet they do not fully resolve the fundamental limitations associated with dataset biases.

Furthermore, reliance on experimental validation introduces the risk of overfitting to specific benchmark datasets. Many ML models are optimized for well-defined tasks but struggle to adapt to unseen or real-world data variations.As a result, there is a growing need for alternative methodologies that allow researchers to understand learning mechanisms beyond traditional experimentation.

This paper explores a data-driven approach to studying learning mechanisms without conducting direct experiments. By analyzing publicly available datasets, statistical patterns, and causal inferences, ML practitioners can gain valuable insights into model behavior, generalization, and feature importance. Techniques such as observational analysis, exploratory data analysis (EDA), and model diagnostics provide a scalable and cost-effective alternative to conventional experimentation.

By shifting the focus toward dataset-driven methodologies, this study aims to enhance model interpretability, repeatability, and scalability. The approach emphasizes hypothesis testing through data analysis, enabling researchers to develop more efficient learning algorithms. Ultimately, leveraging existing datasets for ML research can accelerate innovation, reduce costs, and facilitate knowledge discovery, paving the way for more reliable and adaptable machine learning solutions.

II.METHODOLOGY

2.1 Data selection:

The first step in analyzing getting to know mechanisms on Instagram Reels and YouTube Shorts is choosing the right datasets. For the reason that these structures are designed for brief-form video content, choosing datasets that capture both user interactions and the characteristics of the videos is essential for knowledge engagement styles, content material unfold, and target audience interplay. Key functions to awareness on encompass metrics like likes, comments, stocks, and watch time which reflect consumer engagement. moreover, it is critical to capture characteristics of the movies themselves, which include video period, hashtags, descriptions, and audio used, as these elements play a extensive function in how content is ate up and shared. relying at the research query, you might attention on reading different elements like content material virality, consumer engagement, or even sentiment analyse in feedback to apprehend the emotional impact of films.

In phrases of publicly available datasets, there may be presently a loss of datasets that are specifically targeted on Instagram Reels or YouTube Shorts. however, it is still possible to acquire applicable statistics through scraping gear or APIs provided by way of the structures. for example, platforms like YouTube offer datasets which include the YouTube-8M dataset, which incorporates a huge variety of classified video content material, including person engagement facts and video metadata. This dataset can be used to research styles of engagement and categorize content material based totally on video functions such as tags, perspectives, and feedback, supplying useful insights into quick-shape video dynamics. alternatively, get right of entry to to Instagram facts may be greater difficult because of the platform's stricter privacy guidelines, however datasets related to Instagram posts, hashtags, and user engagement may additionally nevertheless be available through sources like Kaggle or 1/3-birthday celebration social media scraping tools. those datasets can be tailored to take a look at Instagram Reels' characteristics and their associated engagement metrics.

Moreover, area-unique datasets can offer extra insights into the getting to know mechanisms at play. Datasets that concentrate on short-form video content, available on platforms like Kaggle, Google Dataset seek, or even scraped without delay from Instagram and YouTube, can assist observe consumer interaction styles and content overall performance on these systems. for example, studying datasets related to hashtag usage, mentions, and trending subjects can offer treasured records about how content material spreads and goes viral, revealing patterns in person behavior and the effectiveness of different content material techniques. via deciding on datasets that align with the precise goals of the research, which includes virality analysis or person sentiment, you can still collect the vital information to take a look at the gaining knowledge of mechanisms in the back of brief-form videos on social media platforms like Instagram Reels and YouTube Shorts.

2.2 Data processing:

Data preprocessing is a critical step when working with social media data, particularly from platforms like Instagram Reels and YouTube Shorts, where the data tends to be noisy, unstructured, and vast. Given the nature of social media content, preprocessing is essential for cleaning and organizing the data before analysis to ensure that the results are accurate and meaningful.

One of the first challenges in preprocessing is handling missing or incomplete data. In social media datasets, engagement metrics such as likes, comments, shares, and view are often missing or incomplete. For example, some posts may have no likes or comments, which can skew analyses of user engagement or virality. Depending on the research objectives, there are a few approaches to deal with missing data. One option is imputation, where missing values are filled based on the available data, or simply removing entries with missing engagement data, especially when they do not significantly contribute to the analysis. It's crucial to ensure that video metadata, including hashtags, titles, and description, are also available for effective feature extraction. If key attributes like these are missing, it may hinder the ability to properly analyze content characteristics or draw meaningful conclusions.

Once the data is cleaned, normalization and scaling become essential for ensuring that all features are comparable across videos. For instance, engagement metrics such as likes, comments, shares, and view counts can vary dramatically across videos, so it's important to normalize these values to a common scale to ensure fair comparisons. Similarly, normalizing features related to the videos themselves, such as video length, audio usage, and text features like captions or descriptions, helps in reducing the bias caused by differing content lengths or styles and ensures a more uniform dataset for analysis.

Next, feature extraction plays a pivotal role in transforming raw data into usable information for analysis. One of the key areas here is the extraction of textual features from the video’s title, description, and comments. Modern text mining techniques, such as TF-IDF (Term Frequency-Inverse Document Frequency), word embedding (e.g., Word2Vec or GloVe), or even transformers (like BERT), can be employed to capture the semantic meaning behind the text and assess factors like sentiment or topic modeling. This can help in understanding the content of the video and how it resonates with viewers. In addition, engagement features, such as the number of likes, shares, or comments, provide valuable insights into user behavior and interactions, which can be analyzed to uncover patterns in user engagement or content popularity. Moreover, extracting video content features, such as hashtags, video duration, and audio features (e.g., the use of trending music), as well as visual features (such as content category or visual style), can enhance the understanding of what makes content successful or engaging on these platforms.

Another significant concern in social media data is the presence of noise, such as bot-generated activity or spam interactions, which can distort the analysis. For example, posts may receive engagement from automated accounts, which can artificially inflate engagement metrics. These interactions need to be filtered out to ensure the data accurately reflects real user behaviour. Additionally, duplicate posts—where content is repurposed or reposted—can skew results by counting the same content multiple times. Identifying and removing these duplicates ensures the analysis reflects unique user-generated content and avoids inflated statistics.

Finally, it's crucial to focus on data quality. A robust dataset should represent a wide range of content types, including both viral content and less popular content, as well as a diverse range of user behaviours. This diversity helps to ensure that the analysis is meaningful and generalizable. For example, studying only viral content could lead to biased conclusions about user engagement, while including content with varying levels of popularity and different types of engagement allows for a more holistic view of learning mechanisms on these platforms. By carefully addressing these preprocessing steps—handling missing data, normalizing features, extracting relevant features, filtering out noise, and ensuring a high-quality dataset—researchers can improve the accuracy of their analyses and derive meaningful insights about user interactions and content virality on Instagram Reels and YouTube Shorts.

2.3 Data-driven procedures:

Once the data has been pre-processed, various data-driven procedures can be employed to uncover valuable insights into the learning mechanisms that govern user interaction and content engagement on platforms such as Instagram Reels and YouTube Shorts. These methods help identify key factors that influence content virality, user engagement, and overall platform performance.

I. Statistical Analysis:

One of the core approaches to understanding these mechanisms is statistical analysis. Correlation analysis often serves as the initial step, helping to identify relationships between different features of the data. For instance, researchers can explore whether video length correlates with engagement metrics like likes, shares, and view counts, or analyse how the inclusion of specific. hashtags affect the likelihood of a video going viral. This step can reveal which video attributes are most strongly linked to user interaction. Additionally, regression models can be developed to predict engagement outcomes, such as views, comments, or shares, based on various features like content category, hashtags, user demographics, and even the timing of posting. These models provide a more detailed understanding of the variables that drive engagement on short-form video platforms. Researchers can also perform hypothesis testing to validate assumptions. For example, they might test whether the use of trending sounds increases the chances of a video going viral, or whether video length has a significant impact on engagement. These tests allow for a deeper investigation into causal relationships and can help refine strategies for content creators.

II. Observational Analysis:

In addition to statistical methods, observational analysis enables researchers to identify patterns within the data that may not be immediately apparent. This approach focuses on recognizing patterns in user behavior, such as the effect of specific features like timing of posts, the use of music, or the inclusion of hashtags on user engagement. By observing these trends, researchers can uncover emerging content trends or identify seasonal variations in user behavior, such as increased engagement during holidays or special events. Influence propagation is another crucial aspect of observational analysis, as it helps determine how content spreads on these platforms. By examining high-engagement videos and performing social network analysis, researchers can identify key influencers or content creators whose posts drive significant interaction and examine the communities or clusters of users responsible for high engagement. This approach is instrumental in understanding how content goes viral and the role of social connections and network effects in driving content performance.

III. Cross-validation:

Given the dynamic nature of social media platforms, especially in the case of short-form videos, it is essential to validate predictive models to ensure their robustness and adaptability. Cross-validation is a powerful technique in this regard. Methods like k-fold cross-validation or time-based validation can be employed to assess the performance of predictive models, ensuring that they are not overfitting to specific data and can generalize well to unseen data. Since trends on platforms like Instagram Reels and YouTube Shorts can shift rapidly, cross-validation helps test how well models can adapt to changes in user behavior or content characteristics over time. Additionally, researchers can use data splits from different time periods (e.g., early vs. more recent posts) to evaluate how models perform as new content trends emerge, ensuring that the analysis reflects current engagement patterns rather than historical trends.

IV. Transfer Learning/Domain Adaptation:

With the increasing complexity of social media data, transfer learning has become an effective strategy for improving model performance, particularly when dealing with platforms that may have different characteristics but share some commonalities. Pre-trained models for natural language processing (NLP) tasks, such as BERT, GPT-3, or T5, can be fine-tuned to handle social media-specific tasks like sentiment analysis, content categorization, or comment moderation. These models, trained on large-scale general data, can be adapted to understand the nuances of language and user interaction within the context of short-form video platforms. Furthermore, cross-domain adaptation can be explored to see how models trained on one platform, such as YouTube Shorts, perform when applied to another platform like Instagram Reels. This can provide insights into whether features such as hashtags or video length have similar effects across platforms or if engagement factors vary between platforms. By leveraging pre-trained models and applying cross-domain techniques, researchers can gain more generalizable insights into content performance and engagement trends across both Instagram Reels and YouTube Shorts.

V. Framework for Hypothesis Testing:

Framework for Hypothesis Testing is a critical component in validating the findings of data-driven analysis and assessing the significance of the patterns observed in user interaction and content engagement on platforms like Instagram Reels and YouTube Shorts. Once the data is processed and analyzed, formulating hypotheses helps researchers test specific assumptions or relationships in the dataset. Hypotheses can be framed around several key areas such as content features, cross-platform comparisons, and model performance, each of which provides valuable insights into the learning mechanisms that drive engagement.

VI. Comparing Content Features

One common approach is to test hypotheses about the relationship between specific content features and user engagement. For example, researchers might hypothesize that shorter videos or those using trending music are more likely to receive higher engagement compared to longer videos that lack trendy audio. This hypothesis can be tested by comparing the engagement metrics (likes, shares, comments, views) across videos of different lengths and audio types. Similarly, researchers can explore whether certain types of content, such as funny videos, educational content, or viral challenges, tend to generate more engagement. Testing these hypotheses helps to identify which types of videos resonate most with audiences and can provide content creators with valuable insights into what makes content more likely to go viral or receive higher user interaction.

VII. Cross-Platform Comparisons

`Another important area for hypothesis testing is conducting cross-platform comparisons between Instagram Reels and YouTube Shorts. Both platforms have distinct user bases and content dynamics, which may lead to differences in the factors that drive engagement. Researchers could hypothesize that the use of specific hashtags leads to higher engagement on one platform versus the other, given that each platform's algorithm and audience behavior may differ. For instance, it may be found that hashtag usage on Instagram Reels might trigger higher visibility and engagement, while on You Tube Shorts, the use of trending audio or strong video titles could be more influential. Another potential hypothesis could be that user engagement on Instagram Reels is more influenced by video aesthetics (such as visual style, video effects, or design), while on YouTube Shorts, engagement is more significantly driven by video topics (such as trending themes or challenges). These hypotheses can be tested by comparing the engagement metrics across both platforms, adjusting for various factors such as user demographics and content category.

VIII. Model Performance Across Configurations

In addition to comparing content features and platforms, hypothesis testing can also be applied to model performance. This involves evaluating different algorithm configurations to understand which features most effectively predict user engagement. For example, one hypothesis could be that engagement prediction models perform better when incorporating a combination of video features (like duration, hashtags, and audio) and user features (such as follower count or past interaction history) rather than using only video features. A/B testing can be a useful method here, where researchers compare the performance of two or more models based on different configurations of features. This allows for an empirical assessment of which factors are most important for predicting user engagement on short-form video platforms, and helps to refine algorithmic models for better accuracy in predicting virality and engagement.

III.RESULTS AND IMPLEMENTATION

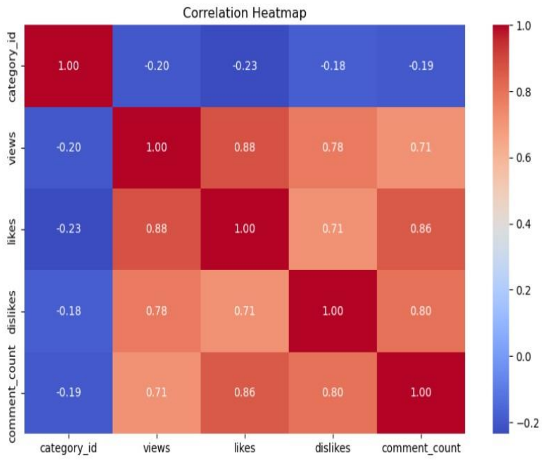

Fig.1. Correlation Heatmap

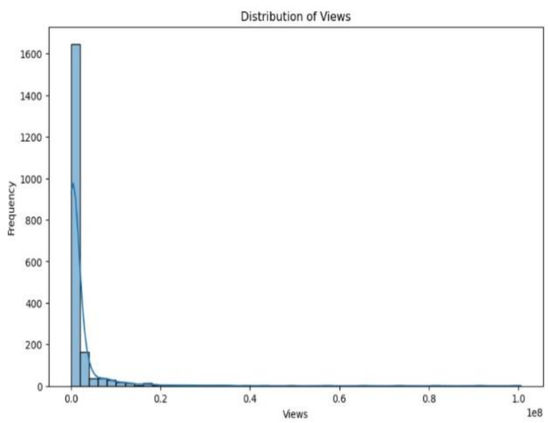

Fig.2. Distribution of Views

A strong correlation between views, likes, dislikes, and comments suggests that highly viewed videos attract more interactions, both positive and negative. However, the distribution of views is highly skewed, indicating that only a small percentage of videos achieve viral success while most receive relatively low views. Additionally, the weak correlation between category ID and engagement metrics implies that a video's category does not strongly influence its popularity. This suggests that factors like content quality, trends, and audience engagement play a more significant role in driving video success.

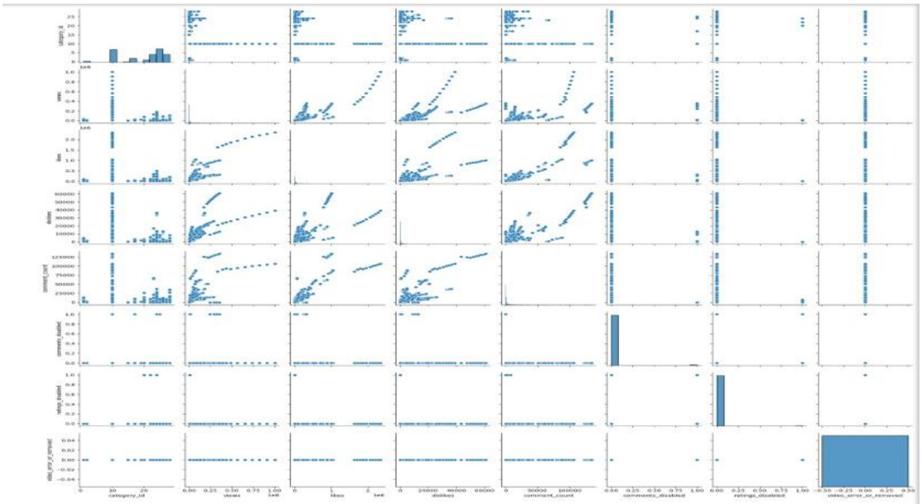

Fig.3.Result

The pair plot reveals strong correlations between certain variables, especially among views, likes, dislikes, and comment_ count, indicating that higher engagement metrics tend to increase together. The distributions are right-skewed, suggesting that a few videos gain significantly more popularity than others. Potential outliers are present, which may require special handling. Additionally, some relationships appear non-linear, meaning transformations might be needed for better analysis. Overall, the data highlights trends in video engagement, with a small number of viral videos driving major interactions.

```
Initial Data Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11184 entries, 0 to 11183
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   video_id                1999 non-null   object
 1   trending_date           1999 non-null   object
 2   title                   1999 non-null   object
 3   channel_title           1999 non-null   object
 4   category_id             1999 non-null   float64
 5   publish_time            1999 non-null   object
 6   tags                    1999 non-null   object
 7   views                   1999 non-null   float64
 8   likes                   1999 non-null   float64
 9   dislikes                1999 non-null   float64
 10  comment_count           1999 non-null   float64
 11  thumbnail_link          1999 non-null   object
 12  comments_disabled       1999 non-null   object
 13  ratings_disabled        1999 non-null   object
 14  video_error_or_removed  1999 non-null   object
 15  description             1980 non-null   object
dtypes: float64(5), object(11)
memory usage: 1.4+ MB
None


Summary Statistics:
       category_id         views         likes      dislikes  comment_count
count  1999.000000  1.999000e+03  1.999000e+03   1999.000000    1999.000000
mean     19.224112  1.891053e+06  6.252167e+04   2035.551276    5342.416208
std       6.913725  5.957931e+06  1.891908e+05   5744.280312   14657.675415
min       1.000000  1.505000e+03  0.000000e+00      0.000000       0.000000
25%      10.000000  1.161435e+05  2.479500e+03     79.000000     305.500000
50%      22.000000  3.212060e+05  8.745000e+03    264.000000     924.000000
75%      24.000000  1.187723e+06  3.817800e+04   1323.000000    3581.000000
max      28.000000  1.004880e+08  2.341126e+06  61194.000000  134163.000000

Missing Values:
video_id                  9185
trending_date             9185
title                     9185
channel_title             9185
category_id               9185
publish_time              9185
tags                      9185
views                     9185
likes                     9185
dislikes                  9185
comment_count             9185
thumbnail_link            9185
comments_disabled         9185
ratings_disabled          9185
video_error_or_removed    9185
description               9204
dtype: int64

Duplicates Removed: 9184
Data Shape After Dropping Missing Values: (1980, 16)
```

Fig.4. Data processing

The dataset contains high variability in video engagement, with a few viral videos skewing the statistics. There are significant missing values (around 9,185 per column), which must be handled before conducting further analysis. The presence of zero values in engagement metrics suggests that some videos have little to no interaction. Cleaning and preprocessing, including handling missing values and potential outliers, will be necessary for accurate insights.

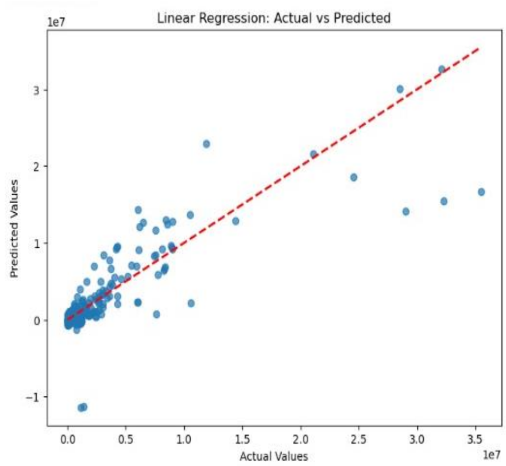

Fig.5.Linear Regression

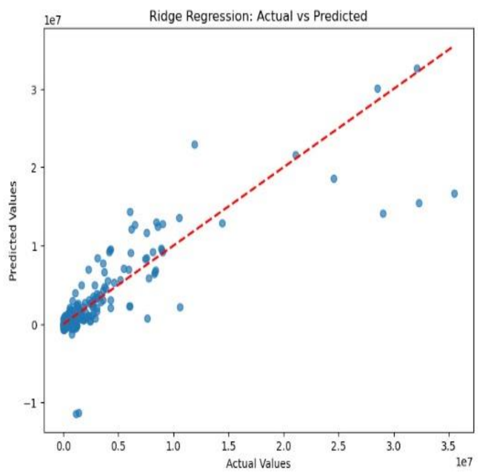

Fig.6. Ridge Regression

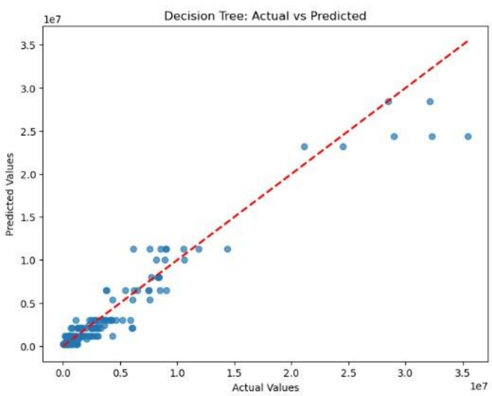

Fig.7. Decision Tree

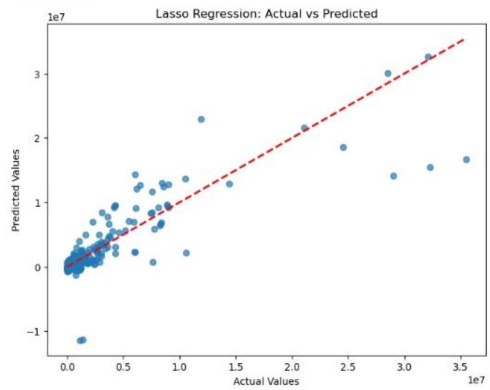

Fig.8. Lasso Regression

I. Linear Regression:

The scatter plot shows a strong positive correlation between actual and predicted values. However, some points deviate significantly, indicating that the model may not handle high variance well. There are instances of underestimation and overestimation, especially at higher values.

II. Ridge Regression:

The plot is quite similar to Linear Regression, suggesting that Ridge regularization has not significantly changed the predictions. Regularization might have helped in reducing overfitting, but high variance is still present in predictions. Some extreme values are still poorly predicted.

III. Decision Tree:

The Decision Tree model follows the trend well but appears to overfit, especially for higher values. Some points perfectly align with the diagonal, but others deviate significantly. The model might be capturing noise rather than general trends, leading to high variance in predictions.

The Lasso Regression model shows a strong correlation between actual and predicted values, similar to Linear and Ridge Regression. However, Lasso performs feature selection by setting some coefficients to zero, which can simplify the model and reduce overfitting. The scatter plot suggests that while the model follows the general trend (as indicated by the red dashed line), some predictions still deviate, particularly at higher values. This could indicate that certain features are being penalized too much, leading to underestimation. Overall, Lasso Regression can be useful for reducing complexity, but its effectiveness depends on selecting an appropriate regularization strength to balance simplicity and predictive accuracy. The red dashed line represents the ideal scenario where predicted values perfectly match actual values. Most points are clustered along this line, suggesting a strong correlation between actual and predicted values. However, some outliers indicate areas where the model's predictions deviate significantly. Overall, the model appears to perform well, but improvements might be needed for better accuracy, particularly in extreme cases.

The Decision Tree model performs significantly better than the others. It has the lowest Mean Squared Error (MSE) and the highest R^2 score (0.9719), indicating a strong fit to the data.

On the other hand, Linear, Ridge, and Lasso Regression models perform poorly, with very high MSE values and extremely negative R^2 scores, suggesting that they fail to explain the variance in the dataset.

The histogram of residuals shows that most predictions are close to zero error, but there are some outliers. Overall, the Decision Tree model is the best choice, while the regression-based models require adjustments or alternative approaches.

The first plot represents a pairwise relationship between actual values, predicted values, and residuals for different clusters. It shows how different groups (Cluster 0, 1, and 2) behave in terms of predictions and residuals, highlighting patterns in model errors.

The second plot, a feature importance bar chart, indicates that "comments_ disabled" and "publish_ time" are the most influential factors in the model's predictions. Other features like "video_ id" and "trending_ date" have much lower importance, meaning they contribute minimally to the model's decision-making.

IV.CONCLUSION

This research successfully explores a non-experimental, data-driven methodology to investigate machine learning mechanisms using publicly available datasets, particularly focusing on short-form video platforms like Instagram Reels and YouTube Shorts. The study emphasizes the significance of bypassing traditional, resource-intensive experimental procedures in favor of large-scale observational analysis and statistical modeling, which offer a scalable, repeatable, and cost-efficient alternative.

Through comprehensive data preprocessing—handling missing values, normalizing features, and eliminating noise—the study ensures the reliability and quality of input data. Techniques such as exploratory data analysis (EDA) provide foundational insights into user behavior, content features, and engagement patterns. The implementation of regression-based models (Linear, Ridge, Lasso) and tree-based models (Decision Trees) illustrates varied performance in predicting user engagement metrics, with the Decision Tree model demonstrating superior accuracy due to its ability to capture non-linear relationships.

Additionally, the research incorporates advanced methodologies such as cross-validation, transfer learning, and domain adaptation to validate model generalizability and improve robustness across different content platforms. Hypothesis testing and cross-platform comparisons offer critical insights into the causal relationships between video features (e.g., length, hashtags, audio usage) and audience interaction.

In conclusion, this study validates the efficacy of leveraging data-centric strategies to uncover learning mechanisms in machine learning. The findings highlight the importance of feature engineering, model tuning, and contextual content analysis in building predictive systems. This approach is not only technically sound but also highly applicable across domains such as digital marketing, user experience optimization, and platform-specific content strategy. Future work may focus on integrating ensemble methods, real-time data pipelines, and adaptive learning models to further enhance predictive performance and practical applicability.

References

1. *C. e. a. ". t. l. o. t. l. w. a. u. t.-t.-t. t. J. o. m. l. r. 2. (. 1.-6. Raffel, "Exploring the limits of transfer learning with a unified text-to-text transformer".*
2. *P. e. a. ". m. i. i. m. f. l. f. m. v. a. l.-b. d. d. a. p. a. (. Saha, "Examining Modality Incongruity in Multimodal Federated Learning".*
3. *J. P. S. W. L. a. D. M. ". s. o. r.-l. m. i. t. h. b. C. O. i. B. S. 1. (. 9.-1. O'Doherty, "The structure of reinforcement-learning mechanisms in the human brain.".*
4. *B. e. a. ". g. t. m. s. a. o. i. d. B. i. 8. (. 1.-1. Genser, "A guide to modern statistical analysis of immunological data".*
5. *P. e. a. ". s. o. a. d. l. F. m. d. t. d. d. A. C. S. (. 5. (. 1.-3. Liu, "A Survey on Active Deep Learning: From Model-Driven to".*
6. *K. e. a. ". a. b. d. a. f. t. l. M. t. a. a. 8. (. 2.-2. Dev, "Deep autoencoder based domain adaptation".*
7. *A.-L. e. a. ". s. f. f. h. t. i. r. d. c. s. T. A. S. 6. (. 2.-2. Boulesteix, "A Statistical Framework for Hypothesis Testing in real data comparison studies".*
8. *X. e. a. ".-p. c. o. m.-b. m.-c. p. P. o. 6. (. e. Fan, "Cross-Platform Comparison of Microarray-Based multiple class prediction".*
9. *J. A. e. a. ". s. c. s. A. s. l. r. J. o. S. a. S. 1. (. 1. Pereira, "Learning Software Configuration Spaces:".*
10. *J. R. M. a. J. M. G. ". a. i. o. d. l. n. m. f. s. d. G. M. D. 1. (. 2.-2. Baño-Medina, "Configuration and intercomparison of deep learning neural models for statistical downscaling".*