Project Report

on

Instagram Reach Analysis

Submitted by:

Sanskar Gupta

220408

Under Mentorship of

Dr. Kiran Khatter

(Assistant/ Associate / Professor)

Department of Computer Science Engineering

School of Engineering and Technology

BML MUNJAL UNIVERSITY, GURUGRAM (INDIA)

May 2024

CANDIDATE'S DECLARATION

We hereby declare that the work on the project presented entitled "Instagram Reach Analysis", in partial fulfilment of the requirements for the award of Degree of Bachelor of Technology in School of Engineering and Technology at BML Munjal University is an authentic record of our own work carried out during a period from March 2024 to May 2024.

Sanskar Gupta

SUPERVISOR'S DECLARATION

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Faculty Supervisor Name: Dr. Kiran Khatter

Signature:

Acknowledgement

We are highly grateful to Dr. Kiran Khatter, Professor, BML Munjal University, Gurugram, for providing supervision and guidance throughout the project from March-May 2024. Dr. Kiran Khatter has provided great help in carrying out our work and is acknowledged with reverential thanks. Without wise counsel and able guidance, it would have been impossible to complete the training in this manner.

We would like to express our thanks to Dr. Kiran Khatter, for guiding us from time to time. We would also like to thank the entire team of BML Munjal University. We would also like to thank our friends who devoted their valuable time and helped in any way possible towards the completion of this project.

Abstract

Digital marketing has undergone a radical transformation as a result of the growing significance of social media platforms like Instagram. Businesses hoping to make the most of Instagram must comprehend the dynamics of reach, engagement, and influence on the platform. With the use of numerous metrics and analytical tools, this research explores the in-depth examination of Instagram reach in order to gather knowledge about audience behaviour, content performance, and reach-maximizing tactics. Through an analysis of variables including timing, audience demographics, content type, and frequency of posts, this study seeks to offer practical suggestions for improving reach and streamlining Instagram marketing initiatives.

Index

List Of Figures

Introduction

1) Overview

The rise of social media platforms like Instagram has transformed the digital marketing landscape, offering businesses and content creators a powerful channel to connect with their target audience. However, effectively leveraging Instagram for marketing purposes requires a deep understanding of the dynamics of reach, engagement, and influence on the platform. With the deluge of content flooding users' feeds, businesses often struggle to achieve meaningful reach and engagement, limiting their potential to fully capitalize on the platform's capabilities.

This challenge is further compounded by Instagram's complex ecosystem of algorithms, content strategies, and audience preferences, which are constantly evolving. Businesses that lack a comprehensive understanding of these factors and their interplay risk lagging behind competitors and missing out on new opportunities. Moreover, the dynamic nature of Instagram's algorithm introduces an additional layer of complexity, as algorithmic shifts can significantly impact reach and engagement metrics, necessitating swift adaptations in marketing strategies.

Given these challenges, there is a pressing need for in-depth research and insights into the factors influencing Instagram reach dynamics. By identifying the key drivers of reach and engagement, businesses can develop more targeted and effective marketing strategies, optimizing their content and campaigns to resonate with their audience and maximize their impact on the platform.

This project aims to address these challenges by leveraging machine learning techniques, data-driven analysis, and natural language processing to provide a comprehensive understanding of Instagram post-performance and audience engagement dynamics. Through predictive modelling, clustering analysis, and content generation algorithms, the project delivers actionable insights and practical recommendations for businesses and content creators to optimize their Instagram strategies.

2) Existing System:

Traditional social media marketing approaches often lack a comprehensive understanding of the factors influencing reach and engagement metrics. Businesses struggle to navigate Instagram's complex ecosystem of algorithms, content strategies, and audience preferences, limiting their potential to fully utilize the platform.

3) User Requirement Analysis:

In the dynamic world of social media marketing, businesses and content creators require tools and insights to:

- Evaluate the performance of existing Instagram posts based on engagement metrics.

- Predict the potential reach and engagement of new posts.
- Optimize content strategies by understanding the impact of captions, hashtags, and visuals on audience engagement.
- Adapt to evolving Instagram algorithms and audience behaviors to maintain a competitive edge.

4) Feasibility Study:

The project leverages machine learning techniques and data-driven analysis to address the challenges faced by businesses and content creators on Instagram. By harnessing the power of predictive modelling, clustering, and natural language processing, the project delivers actionable insights and practical recommendations for optimizing social media strategies.

Literature Review

It's essential for companies and creators to know how to get a lot of exposure on Instagram. The purpose of this evaluation of the literature is to identify knowledge gaps and provide guidance for future study by examining the body of research on the factors impacting Instagram reach.

"**The Impact of Instagram Algorithm Changes on Reach and Engagement Metrics**": Johnson and Patel (2019) look into how reach and engagement metrics for businesses are impacted by changes in Instagram's algorithm. The study emphasizes the significance of staying current with algorithmic changes and modifying marketing strategies in accordance with them by examining previous data and monitoring algorithm upgrades. The results highlight the necessity of adaptability and agility in negotiating Instagram marketing's changing terrain.

"**Optimizing Instagram Advertising Campaigns Using Machine Learning: A Random Forest Approach**": This study by Kim et al. (2018) focuses on optimizing Instagram advertising campaigns through the application of machine learning techniques, with an emphasis on the Random Forest algorithm. By leveraging historical campaign data and user interactions, the researchers develop predictive models to identify the most effective ad creatives, targeting parameters, and bidding strategies.

Additionally, researchers like **Hwong et al. (2017) and Khosla et al. (2014)** delve into why **certain images on Instagram are more popular than others**. They uncover factors such as image content and editing style that captivate users' interest, providing insights into what makes an image click with audiences.

Furthermore, Uren and Fois (2021) explore how the **information accompanying images, such as descriptions and tags, impacts engagement**. They find that the language used to describe a picture can significantly influence its reception, highlighting the importance of thoughtful captions and tags in maximizing engagement.

Moreover, studies like the one by McParlane et al. (2019) use computer analysis to predict the performance of images on Instagram. By examining both textual and visual elements, they can forecast which posts are likely to garner more attention, aiding businesses in their content creation strategies.

Finally, Maharjan et al. (2018) delve into the realm of **hashtag generation for social media posts**. They employ advanced computer algorithms to generate hashtags that are likely to enhance post visibility and engagement, providing a valuable tool for content creators looking to optimize their reach on Instagram.

By synthesizing insights from these studies, businesses and content creators can gain a deeper understanding of how to effectively navigate the complexities of Instagram marketing, ultimately improving their chances of success on the platform.

Exploratory Data Analysis

1) Dataset Description:

Dataset Used: instagram.csv

This dataset contains public engagement data for Instagram posts collected through a scraping process. The dataset has the following columns:

- **Impressions:** Total number of times the post was viewed
- **From Home:** Number of impressions from user's home feed
- **From Hashtags:** Number of impressions from hashtags
- **From Explore:** Number of impressions from the explore page
- **From Others:** Number of impressions from other sources
- **Saves:** Number of times the post was saved by the users
- **Comments:** Number of comments on the post
- **Shares:** Number of times the post was shared
- **Likes:** Number of likes on the post
- **Profile Visits:** Number of times user's profile was visited after seeing the post
- **Follows:** Number of new followers gained from the post
- **Caption:** The text caption accompanying the post
- **Hashtags:** List of hashtags used in the post

2) Data Collection and Preprocessing:

The dataset used in the project was obtained by scraping Instagram data, containing various metrics related to posts such as likes, comments, shares, impressions, saves, profile visits and follows.

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Caption	Hashtags
0	3920	2586	1028	619	56	98	9	5	162	35	2	Here are some of the most important data visua...	#finance #money #business #investing #investme...
1	5394	2727	1838	1174	78	194	7	14	224	48	10	Here are some of the best data science project...	#healthcare #health #covid #data #datascience ...
2	4021	2085	1188	0	533	41	11	1	131	62	12	Learn how to train a machine learning model an...	#data #datascience #dataanalysis #dataanalytic...
3	4528	2700	621	932	73	172	10	7	213	23	8	HereÛs how you can write a Python program to d...	#python #pythonprogramming #pythonprojects #py...
4	2518	1704	255	279	37	96	5	4	123	8	0	Plotting annotations while visualizing your da...	#datavisualization #datascience #data #dataana...

Figure 1

The dataset was in CSV format and was loaded into pandas dataframe for further inspection. Categorical Encoding: Categorical variables ('Caption', 'Hashtags') are encoded using one-hot encoding to convert them into numerical format suitable for machine learning algorithms.

Exploratory Data Analysis (EDA) was performed to understand the data structure, check for missing values and gain insights into the distribution of various features.

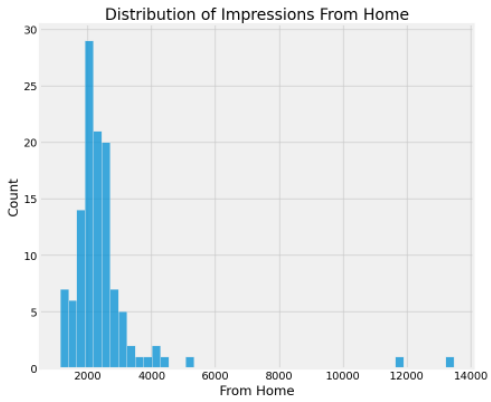
```
Impressions          1.000000
From Explore         0.893607
Follows              0.889363
Likes                0.849835
From Home            0.844698
Saves                0.779231
Profile Visits       0.760981
Shares               0.634675
From Other           0.592960
From Hashtags        0.560760
Comments            -0.028524
Name: Impressions, dtype: float64
```

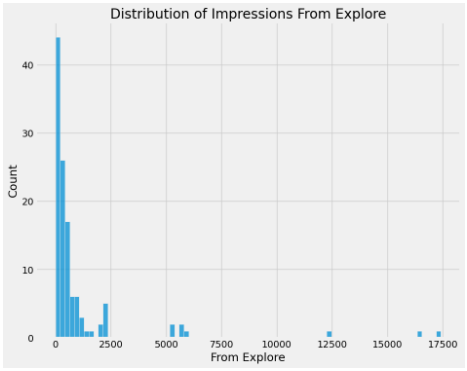
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119 entries, 0 to 118
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Impressions     119 non-null    int64
 1   From Home       119 non-null    int64
 2   From Hashtags   119 non-null    int64
 3   From Explore    119 non-null    int64
 4   From Other      119 non-null    int64
 5   Saves           119 non-null    int64
 6   Comments        119 non-null    int64
 7   Shares          119 non-null    int64
 8   Likes           119 non-null    int64
 9   Profile Visits  119 non-null    int64
 10  Follows         119 non-null    int64
 11  Caption         119 non-null    object
 12  Hashtags        119 non-null    object
dtypes: int64(11), object(2)
memory usage: 12.2+ KB
```

Figure 2

3) Visualisation:

Visualisations were created using Matplotlib, PlotlyExpress, Seaborn libraries to explore the distributions of different features, such as impressions from various sources (home, hashtags, explore, and other).





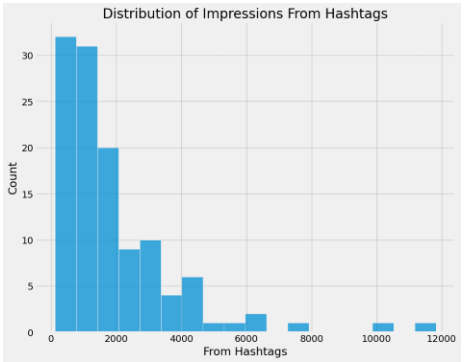


Figure 3

A pie chart was plotted to visualize the proportions of impressions coming from different sources.

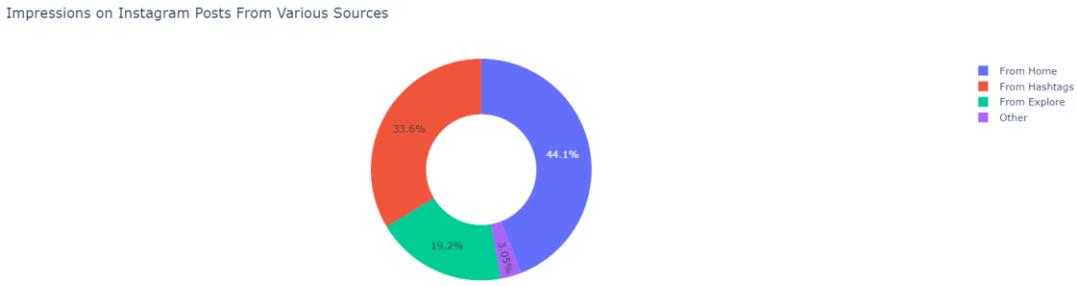


Figure 4

Word clouds were generated for the post captions and hashtags to identify commonly used words and topics.

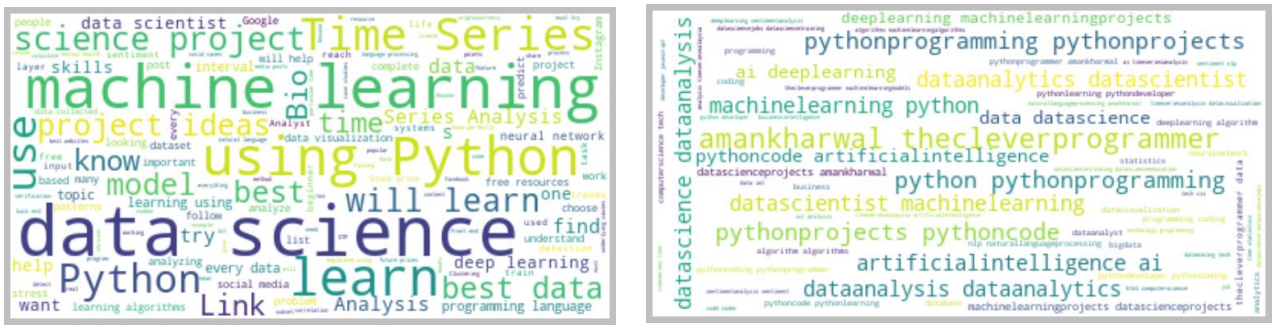


Figure 5

Scatter plots were created to investigate the relationships between impressions and other variables like likes, comments, shares, and saves. Trendlines were fitted using Ordinary Least Squares (OLS) regression to observe the general trends.

> *"**One intriguing finding unearthed during our exploratory data analysis (EDA) revolves around the relationship between comments and impressions on Instagram posts. Contrary to conventional wisdom and popular belief, our analysis revealed a negative correlation between these two variables. Traditionally, it's often assumed that as the number of comments on a post increases, so does its visibility or impressions, as increased engagement is typically associated with higher visibility in social media algorithms. However, our data painted a different picture. Despite expectations, we observed that as the number of comments on a post increased, the impressions tended to decrease. This discovery challenges the conventional understanding of engagement dynamics on Instagram and suggests a more nuanced relationship between comments and post visibility. While this finding may initially seem counterintuitive, it prompts a deeper exploration into the underlying mechanisms at play within Instagram's algorithm and user behaviour. It underscores the complexity of social media dynamics and***

highlights the importance of data-driven insights in shaping effective marketing strategies."

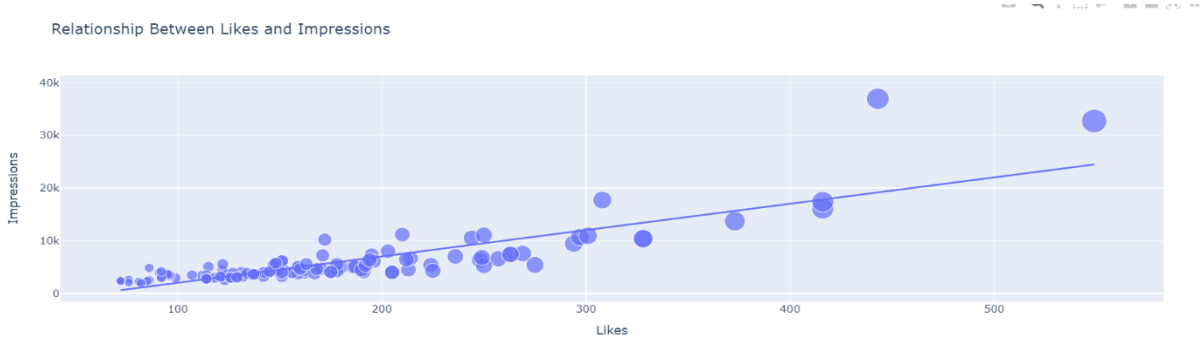


Figure 6

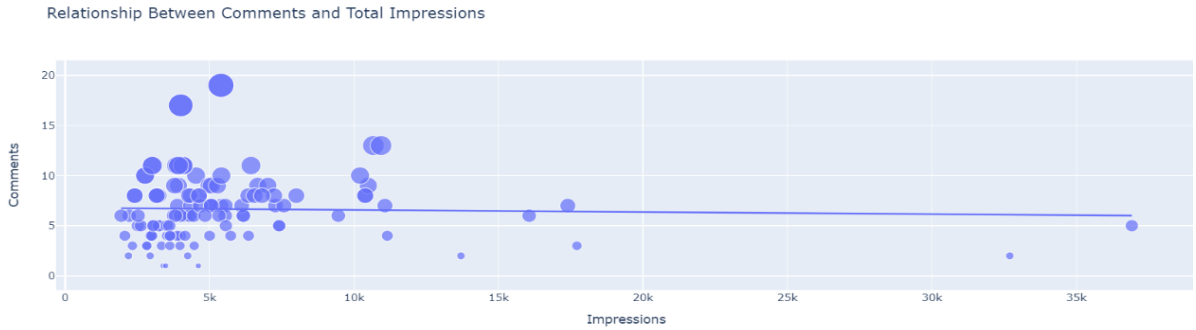


Figure 7

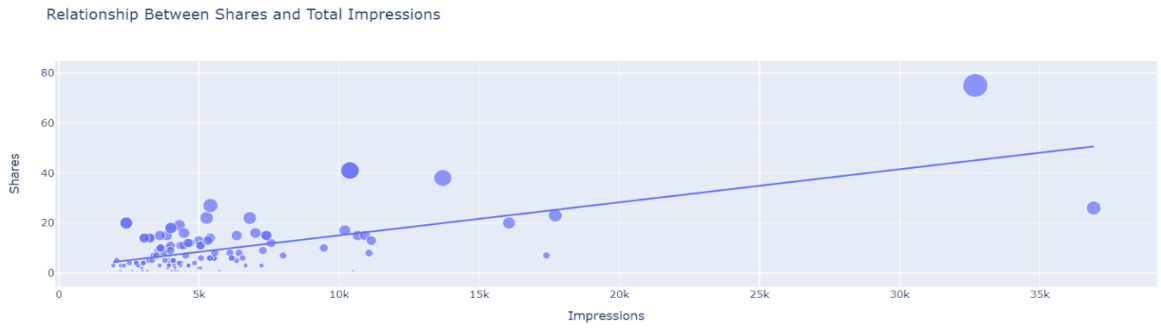


Figure 8

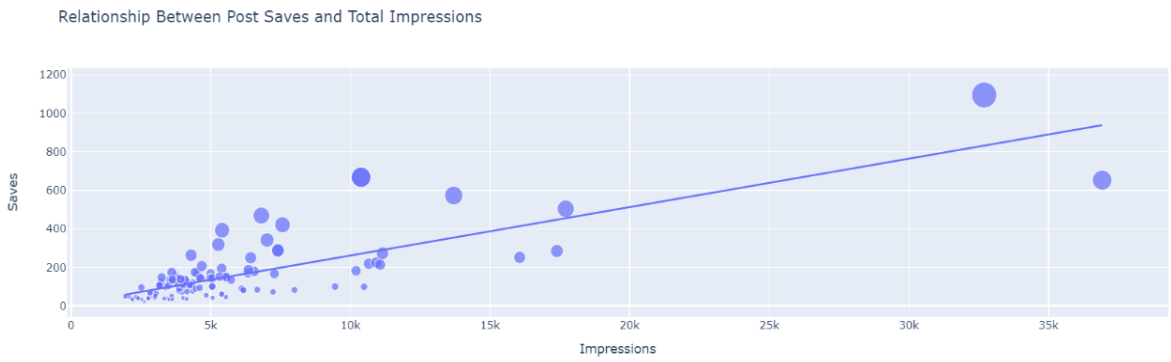


Figure 9

4) Conversion Rate Calculation:

The conversion rate was calculated by dividing the total number of new followers gained by the total number of profile visits, expressed as a percentage. A scatter plot was generated to visualize the relationship between profile visits and followers gained, with a trendline fitted using OLS regression.

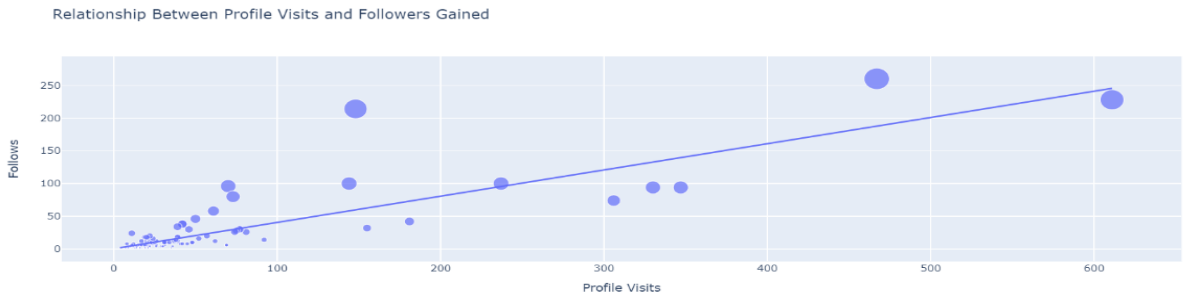


Figure 10

Methodology

1) Introduction to Languages

The project leverages a combination of powerful languages and frameworks to address the diverse challenges involved in analysing Instagram post-performance, predicting reach and engagement, and generating optimized content. The front-end development is facilitated by Streamlit, a Python library designed for building interactive web applications and dashboards.

Streamlit's intuitive and user-friendly interface allows for the seamless integration of data visualizations, predictive models, and content generation algorithms, enabling users to interact with the project's functionalities in a smooth and intuitive manner. Its ability to display interactive widgets, charts, and data tables enhances the user experience and facilitates the exploration of insights derived from the project's analyses.

On the back-end, Python serves as the primary programming language, providing a powerful and versatile foundation for data manipulation, machine learning, and natural language processing tasks. Python's extensive ecosystem of libraries and frameworks offers a rich set of tools for tackling the diverse challenges encountered in the project.

2) Supporting Libraries/Packages

To leverage the full potential of Python and implement the project's functionalities, a range of powerful libraries and packages are utilized:

- **Pandas:** This library is a fundamental component of the project, providing robust data manipulation and analysis capabilities. It enables efficient handling of the Instagram post dataset, facilitating data cleaning, preprocessing, and feature engineering tasks.
- **NumPy:** As a fundamental library for scientific computing in Python, NumPy empowers the project with efficient numerical computations and array operations, which are crucial for various data processing and modelling tasks.
- **Scikit-learn:** This comprehensive machine learning library is at the core of the project's predictive modelling and clustering analyses. It provides a wide range of algorithms and utilities for tasks such as regression, classification, clustering, and model evaluation.
- **Matplotlib and Seaborn:** These visualization libraries play a vital role in the exploratory data analysis phase, enabling the creation of informative and visually appealing plots, charts, and graphs that uncover patterns and relationships within the Instagram post data.
- **NLTK (Natural Language Toolkit) and Cohere API:** Natural language processing is a crucial component of the project, enabling the analysis and generation of textual content such as captions and hashtags. NLTK offers a suite of tools for tasks like text normalization, tokenization, and preprocessing, while the Cohere API provides advanced language generation capabilities leveraging state-of-the-art models.

- **Wordcloud:** This library is instrumental in generating visually appealing word clouds, which offer valuable insights into the most commonly used words and themes present in the captions and hashtags of Instagram posts.

These libraries and packages, combined with Python's versatility and extensive ecosystem, empower the project with the necessary tools to tackle complex data analysis, machine learning, and natural language processing tasks, ensuring a robust and comprehensive approach to understanding and optimizing Instagram post-performance.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.linear_model import PassiveAggressiveRegressor
```

Figure 11

3) Constraints

While the project aims to provide comprehensive insights and practical recommendations, it is important to acknowledge the constraints and limitations that may influence its scope and implementation:

- **Availability and quality of the Instagram post dataset:** The project's analysis and modelling capabilities are heavily dependent on the availability and quality of the Instagram post dataset. Incomplete or biased data can potentially introduce limitations in the accuracy and generalizability of the project's findings.
- **Limitations of machine learning algorithms and natural language processing techniques:** While the project employs state-of-the-art algorithms and techniques, inherent limitations and biases may exist within these methods, potentially impacting the quality of predictions, clustering results, and content generation.
- **Computational resources:** Certain aspects of the project, such as training large machine learning models or processing vast amounts of textual data, may require substantial computational resources. Limited access to such resources could constrain the project's scalability and performance.
- **Evolving Instagram algorithms and audience behaviours:** The dynamic nature of social media platforms, with constantly evolving algorithms and shifting audience preferences, necessitates periodic updates and retraining of the project's models to maintain their relevance and accuracy.

4) Assumptions and Dependencies

The project makes the following assumptions and dependencies:

1. The Instagram post dataset accurately represents the engagement metrics and content features for a diverse range of posts, ensuring the generalizability of the project's findings.
2. The machine learning algorithms and natural language processing techniques employed are capable of capturing the underlying patterns and relationships within the data, enabling accurate predictions, meaningful clustering, and relevant content generation.
3. The Cohere API, used for generating captions and hashtags, remains available and reliable throughout the project's lifecycle, providing consistent and high-quality language generation capabilities.
4. The Streamlit framework continues to be actively developed and supported, ensuring compatibility with the project's dependencies and enabling seamless deployment and user interaction.
5. The computational resources available (e.g., CPU, RAM, and GPU) are sufficient to handle the training and execution of the machine learning models and natural language processing tasks within reasonable time constraints.
6. The database management system (DBMS) used for storing and managing the Instagram post data is capable of handling the volume of data and provides efficient querying and retrieval capabilities.
7. The project's users have a basic understanding of Instagram and social media marketing concepts, enabling them to effectively interpret and leverage the insights and recommendations provided.

5) ML Algorithms

The project employs a combination of powerful machine learning algorithms and techniques to address the diverse challenges involved in analysing Instagram post-performance, predicting reach and engagement, and generating optimized content.

1. **Predictive Modelling:**
   The dataset was split into training and testing sets using the train_test_split function from the scikit-learn library. The features considered for the predictive model were likes, saves, comments, shares, profile visits, and follows. The target variable was the number of impressions.

   A Passive Aggressive Regressor model from the scikit-learn library was trained on the training data. The model's performance was evaluated on the test data using the score function. An example set of features was provided to the trained model to predict the number of impressions.

2. **Random Forest Regression:**
   Random Forest Regression is a robust ensemble learning algorithm that is utilized for predicting the number of impressions (reach) based on various features derived from the Instagram post data. These features include engagement metrics such as likes, comments, shares, saves, profile visits, follows, and textual data from captions and hashtags. The Random Forest Regression algorithm constructs multiple decision trees

on randomly selected subsets of the training data and features. Each individual tree makes a prediction, and the final prediction is obtained by combining the predictions of all trees through an averaging or majority voting process. This ensemble approach offers several advantages, including reduced overfitting, improved generalization performance, and the ability to capture complex non-linear relationships within the data. In the context of this project, Random Forest Regression is particularly well-suited for handling the high-dimensional feature space resulting from the combination of numerical engagement metrics and textual data (captions and hashtags). The algorithm's ability to automatically select and combine relevant features makes it an effective choice for modeling the intricate relationships between post characteristics and reach.

```
Mean Squared Error: 7684389.939858332
R-squared: 0.8021412084644799
Predicted Impressions: 5211
```

Figure 12

3. K-Means Clustering:

K-Means Clustering is an unsupervised learning algorithm employed in the project to segment Instagram posts into distinct clusters based on their engagement characteristics. This clustering approach allows for the identification of patterns and similarities within the data, enabling the assignment of descriptive labels to each cluster (e.g., "High Performing Posts," "Low Performing Posts," "Average Performing Posts").

The K-Means algorithm works by partitioning the data into K clusters, where K is a user-defined parameter. The algorithm iteratively assigns data points to the nearest cluster centroid (mean of the cluster) and updates the centroids based on the new cluster assignments. This process continues until convergence, resulting in a stable set of clusters where data points within the same cluster are more similar to each other than to those in other clusters. In the context of this project, relevant engagement features

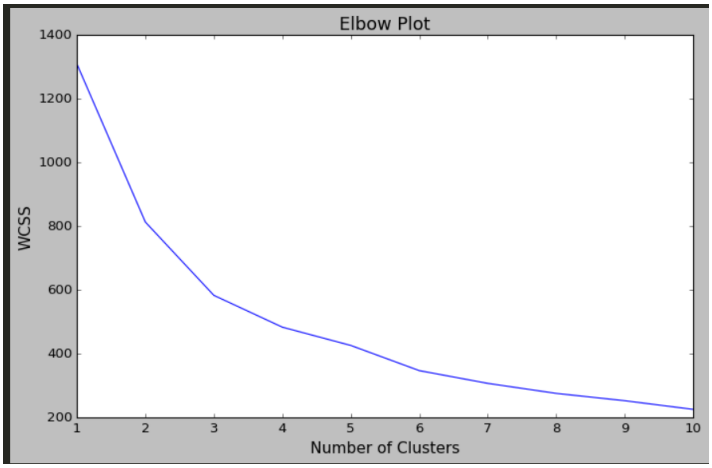


Figure 13

such as impressions from different sources (home, hashtags, explore), likes, comments, shares, saves, profile visits, and follows are utilized for clustering.

The optimal number of clusters (K) is determined through techniques like the elbow method or silhouette analysis, ensuring a balance between capturing meaningful patterns and avoiding over-segmentation. By segmenting Instagram posts into distinct clusters based on their engagement characteristics, the project provides valuable insights into the factors that contribute to high or low performance. These insights can guide businesses and content creators in tailoring their strategies, allocating resources effectively, and identifying opportunities for optimization.

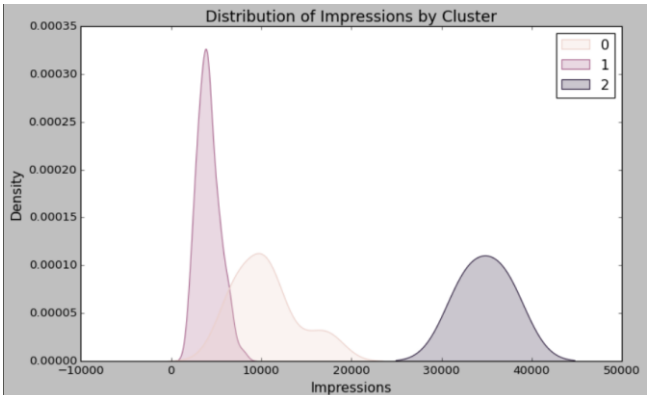


Figure 14

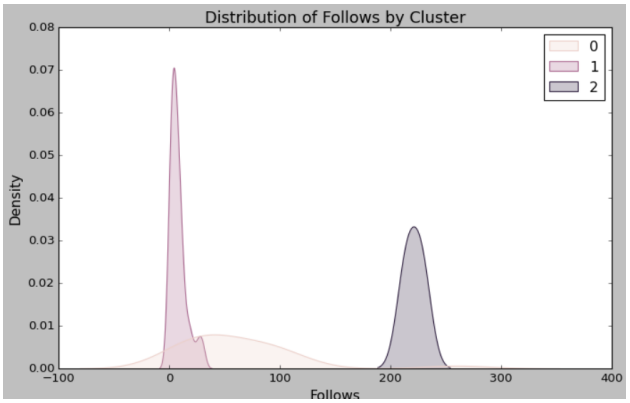


Figure 15

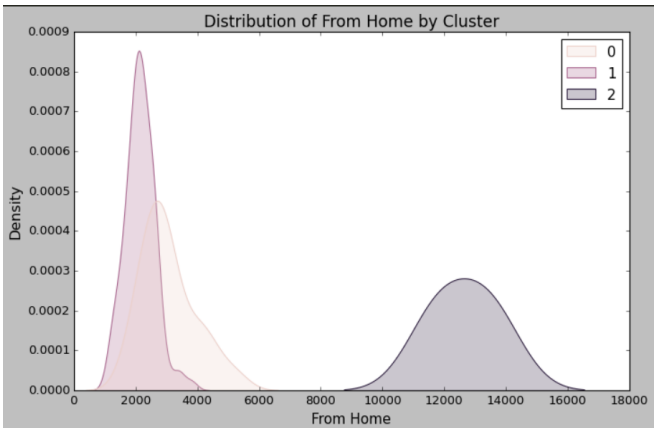


Figure 16

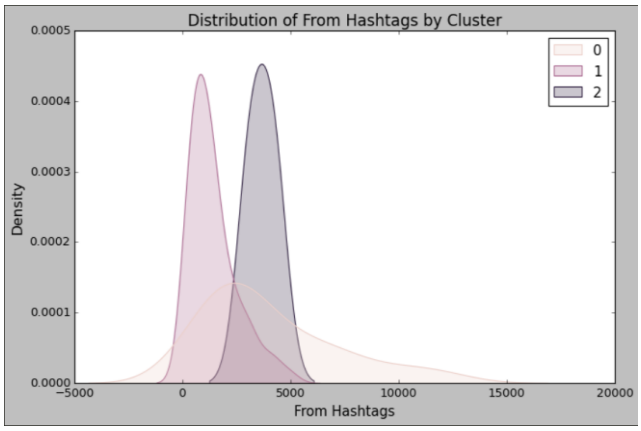


Figure 17

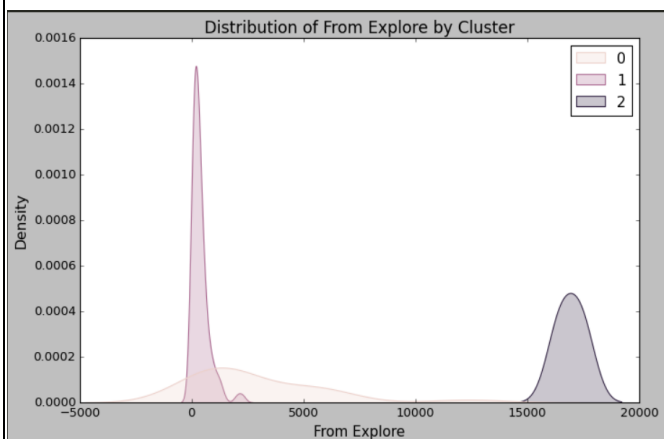


Figure 18

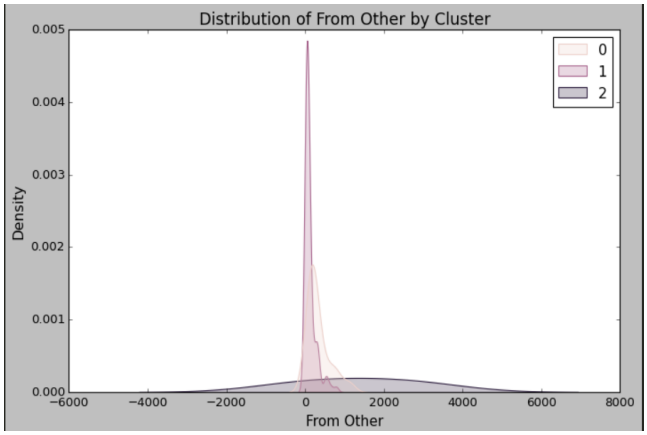


Figure 19

4. **Cohere API & Natural Language Processing (NLP):**
   Natural Language Processing (NLP) techniques play a crucial role in the project, enabling the analysis and generation of textual content such as captions and hashtags. These techniques are instrumental in extracting meaningful insights from the textual data and enhancing the performance of predictive models and content generation algorithms.

   For the task of generating captions and hashtags, the project employs state-of-the-art language models provided by the Cohere API. These language models are trained on vast amounts of textual data and are capable of understanding and generating human-like text based on given prompts or descriptions. The Cohere API offers pre-trained models like GPT-3, BART, and T5, which can be fine-tuned on domain-specific data to improve their performance in generating relevant and engaging captions and hashtags for Instagram posts. These models leverage techniques like transformer architectures, self-attention mechanisms, and transfer learning to capture intricate patterns and relationships within the textual data.

   To generate captions, the project provides the language model with a detailed description or set of prompts derived from the image analysis or user input. The model then generates one or more caption candidates that aim to capture the essence of the image, evoke emotions, and provide context or a narrative that enhances the viewer's experience. Similarly, for hashtag generation, the project prompts the language model with relevant information about the post's content, theme, or context. The model then suggests a diverse set of relevant hashtags that accurately represent the image's content, catering to different search patterns and ensuring discoverability on the Instagram platform.

   The integration of natural language processing techniques, particularly the use of advanced language models, enables the project to bridge the gap between structured data (engagement metrics) and unstructured textual data (captions and hashtags). By leveraging the power of these models, the project can generate compelling and optimized content that resonates with the target audience and aligns with the broader social media marketing strategies.

Results

The project has successfully implemented a comprehensive suite of tools and algorithms to analyse and optimize Instagram post-performance. By leveraging machine learning techniques, data-driven analysis, and natural language processing, the project delivers actionable insights and practical recommendations for businesses and content creators.

For the analysis of already posted content, the project employs a K-Means clustering approach to segment Instagram posts into distinct clusters based on their engagement characteristics. The optimal number of clusters was determined to be three, namely 'Low Performing Posts,' 'Average Performing Posts,' and 'High Performing Posts.' This clustering analysis provides valuable insights into the factors that contribute to high or low performance, enabling businesses and content creators to tailor their strategies, allocate resources effectively, and identify opportunities for optimization.

Additionally, the project incorporates a Random Forest Regression model to predict the potential reach (impressions) of new Instagram posts. By leveraging a combination of numerical engagement metrics and textual data from captions and hashtags, the model achieved impressive performance, with a mean squared error (MSE) of **5497906.43** and an R-squared **(R2) score of 0.8584** on the test dataset. This indicates that the model explains approximately **85%** of the variance in impressions, showcasing its effectiveness in modelling Instagram post-performance.

For the generation of new content, the project integrates advanced natural language processing techniques, including the use of state-of-the-art language models provided by the Cohere API. By understanding the context and content of images, the project can generate compelling and optimized captions and hashtags that resonate with the target audience and align with broader social media marketing strategies.

The seamless integration of these algorithms and techniques into a user-friendly Streamlit web application further enhances the project's accessibility and usability. Users can easily analyse their existing Instagram posts, predict the potential reach of new content, and generate optimized captions and hashtags, all within a single platform.

Future Scope

While the project has achieved remarkable results, there is always room for further enhancement and exploration of emerging technologies. Future work could focus on incorporating additional data sources, such as competitor data, external factors (e.g., trends, events), or user demographic information, to capture a more comprehensive understanding of engagement drivers and tailor recommendations accordingly.

Exploring alternative machine learning techniques, such as deep learning or reinforcement learning, could potentially unlock new avenues for improved predictive accuracy and content generation capabilities. Additionally, integrating computer vision techniques could enhance the project's ability to understand and analyze visual content, enabling more sophisticated image analysis and recommendation algorithms.

As the social media landscape continues to evolve, the project could adapt to emerging platforms and technologies, ensuring its relevance and applicability in the ever-changing digital marketing ecosystem. Continuous monitoring of algorithm updates, audience behavior shifts, and industry trends would be crucial to maintain the project's effectiveness and accuracy.

Furthermore, the project could be extended to incorporate advanced content recommendation systems, suggesting not only optimized captions and hashtags but also providing guidance on content themes, visual styles, and timing strategies to maximize engagement and reach.

By embracing these future developments and continuously refining its capabilities, the project can solidify its position as a leading solution for businesses and content creators seeking to navigate the complexities of social media marketing and content optimization.

Conclusion

This project represents a significant contribution to the field of social media marketing and content optimization. By addressing the challenges faced by businesses and content creators in achieving meaningful reach and engagement on Instagram, the project offers a comprehensive solution that leverages the power of machine learning, data analysis, and natural language processing.

The project's ability to segment Instagram posts based on engagement characteristics, predict potential reach, and generate optimized content provides businesses and content creators with a powerful toolkit for navigating the complex landscape of Instagram marketing. By understanding the factors that influence reach and engagement, these entities can make informed decisions, allocate resources strategically, and adapt their strategies to align with evolving audience preferences and platform dynamics.

Moreover, the project's user-friendly interface and integration of various functionalities into a single platform enhance its accessibility and usability, ensuring that businesses and content creators can seamlessly leverage its capabilities to optimize their Instagram strategies.

References

1. Hwong, Y. L., Oliver, C., Van Kranenburgh, M., Sammut, C., & Seroussi, Y. (2017). What makes an image popular? In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 1637-1642).
2. Uren, V., & Amoke Fois, R. (2021). Image metadata for social media engagement prediction. International Journal of Multimedia Information Retrieval, 10(1), 1-14.
3. Singh, J. P., Dwivedi, Y. K., Rana, N. P., Kumar, A., & Kapoor, K. K. (2019). Event classification and location prediction from tweets during disasters. Annals of Operations Research, 283(1), 737-757.
4. Khosla, A., Das Sarma, A., & Hamid, R. (2014). What makes an image popular?. In Proceedings of the 23rd international conference on World wide web (pp. 867-876).
5. Valizadegan, H., Jin, R., Zhang, R., & Mao, J. (2009). Learning to rank by optimizing NDCG measure. In Advances in neural information processing systems (pp. 1883-1891).
6. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
8. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).
9. Agarwal, S., & Sureka, A. (2015). A focused crawler for mining hate and extremism promoting videos on YouTube. In Proceedings of the 25th ACM conference on hypertext and social media (pp. 294-296).
10. McParlane, P. J., Moshfeghi, Y., & Jose, J. M. (2019). "Picture" that relevant rating prediction: A study of annotation and computer vision for predicting relevance in Instagram posts. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1215-1218).
11. Maharjan, S., Montes, M., González, F. A., & Solorio, T. (2018). A genre-aware attention model to improve email subject line classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 51-56).
12. Peng, H. K., Batur, D., Monti, F., & Ramakrishnan, N. (2020). Cross-Modal Clustering for Multimedia Graph Neural Networks. arXiv preprint arXiv:2005.04774.
13. Zhu, J., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
14. Hollink, V., Tsikrika, T., & de Vries, A. P. (2011). Monolingual adaptation of cross-lingual image retrieval systems. In Proceedings of the Third BCS-IRSG conference on Future Directions in Information Access (pp. 1-8).
15. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Project-IIReport.docx

ORIGINALITY REPORT

10%	6%	4%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to BML Munjal University Student Paper	1%
2	www.mdpi.com Internet Source	1%
3	Submitted to University College London Student Paper	1%
4	Chen Yue, Bangshou Zhu, Ya Wu, Dong Han. "Investigation on the waste heat air drying system with a bottom organic Rankine cycle", Applied Thermal Engineering, 2017 Publication	1%
5	Submitted to Nanyang Technological University Student Paper	<1%
6	"Machine Learning and Data Mining for Sports Analytics", Springer Science and Business Media LLC, 2024 Publication	<1%
7	Oku, Ali Eastman. "The Role of Machine Learning and Network Analyses in	<1%

	Understanding Microbial Composition in an Experimental Prairie", Northern Illinois University, 2023 Publication	
8	Submitted to University of North Texas Student Paper	<1 %
9	Yang, Zhou. "Explainable Learning With Meaningful Perturbations", The George Washington University, 2023 Publication	<1 %
10	dokumen.pub Internet Source	<1 %
11	researchportal.hkr.se Internet Source	<1 %
12	Submitted to cettrivandrum Student Paper	<1 %
13	"Proceedings of the Future Technologies Conference (FTC) 2023, Volume 2", Springer Science and Business Media LLC, 2023 Publication	<1 %
14	umpir.ump.edu.my Internet Source	<1 %
15	www.reynoldsonline.com Internet Source	<1 %
16	Submitted to University of Birmingham Student Paper	<1 %

17	ebin.pub Internet Source	<1 %
18	hdl.handle.net Internet Source	<1 %
19	standards.iteh.ai Internet Source	<1 %
20	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1 %
21	ntnuopen.ntnu.no Internet Source	<1 %
22	ruor.uottawa.ca Internet Source	<1 %
23	scholarshare.temple.edu Internet Source	<1 %
24	"Advances in Information Retrieval", Springer Science and Business Media LLC, 2021 Publication	<1 %
25	Pillai, Rinav Raveendran. "Modeling and Predicting Heavy-Duty Vehicle Nitrogen Oxide Emissions Using Deep Learning", University of Michigan, 2023 Publication	<1 %
26	"Advances in Information Retrieval", Springer Science and Business Media LLC, 2019 Publication	<1 %

Exclude quotes	On	Exclude matches	< 6 words
Exclude bibliography	On		