

A Project Report on

Predicting Instagram Influencer Engagement Rate

Submitted in partial fulfillment of the requirements for the award of
the degree of

Bachelor of Engineering

in

Computer Engineering

by

Aishwarya More(18102008)

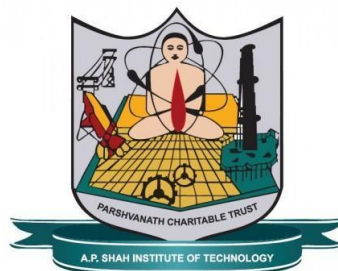
Shweta patil(19202002)

Preksha Vora(19202004)

Moksha shah(18102057)

Under the Guidance of

Prof. Shafaque Fatma Syed



Department of Computer Engineering

NBA Accredited

A.P. Shah Institute of Technology
G.B.Road,Kasarvadavali, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2021-2022

Approval Sheet

This Project Report entitled “*Predicting Instagram Influencer Engagement Rate*” Submitted by “*Aishwarya More*”(18102008), “*Shweta Patil*”(19202002), “*Preksha Vora*”(19202004), “*Moksha Shah*”(18102057) is approved for the partial fulfillment of the requirement for the award of the degree of *Bachelor of Engineering* in *Computer Engineering* from *University of Mumbai* .

(Prof. Shafaque Fatma Syed)

Guide

Prof. Sachin Malave

Head Department of Computer Engineering

Place: A.P. Shah Institute of Technology, Thane

Date:

CERTIFICATE

This is to certify that the project entitled “***Predicting Instagram Influencer Engagement Rate***” submitted by “***Aishwarya More***” (18102008), “***Shweta Patil***” (19202002), “***Preksha Vora***”(19202004), “***Moksha Shah***” (18102057) for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Computer Engineering***, to the University of Mumbai, is a bonafide work carried out during academic year 2021-2022.

(Prof. Shafaque Fatma Syed)

Guide

Prof. Sachin Malave
Head Department of Computer Engineering

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place: A.P. Shah Institute of Technology, Thane

Date:

Acknowledgement

We have great pleasure in presenting the report on **Predicting Instagram Influencer Engagement Rate**. We take this opportunity to express our sincere thanks towards our guide **Shafaque Fatma Syed** Department of CS, APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of the project.

We thank **Prof. Sachin Malave** Head of Department,CS, APSIT for his encouragement during the progress meeting and providing guidelines to write this report.

We thank **Prof. Amol Kalugade** BE project coordinator, Department of CS, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Student Name1: Aishwarya More Student ID1: 18102008

Student Name2: Shweta Patil Student ID2: 19202002

Student Name3: Preksha Vora Student ID3: 19202004

Student Name4: Moksha Shah Student ID4: 18102057

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Aishwarya More - 18102008)

(Shweta Patil - 19202002)

(Preksha Vora - 19202004)

(Moksha Shah - 18102057)

Date:

Contents

1	Project Conception and Initiation	1
1.1	Abstract	1
1.2	Introduction	1
1.3	Objective	2
1.4	Literature review	3
1.5	Problem Definition	4
1.6	Scope	4
1.7	Technology Stack	5
1.8	Benefits for Environment and Society	6
2	Project Design	7
2.1	Proposed System	7
2.2	Design(Flow of Modules)	8
2.3	Description Use Case	9
2.4	Class Diagram	10
2.5	Modules	10
2.5.1	Module 1.	10
2.5.2	Module 2.	10
2.5.3	Module 3.	11
2.5.4	Module 4.	12
3	Implementation	14
3.1	Proposed system-implementation.	14
3.1.1	Policies.	14
3.1.2	Algorithms	15
3.1.3	Pseudo code.	15
3.1.4	Platforms for execution.	16
4	Result.	17
5	Conclusion and Future Scope.	24
6	References	25
7	Bibliography.	26
8	Certificate.	27
9	Log Book	29

List of Figures

Figure	Name of figure	Page no
1.1	Literature Review	04
2.1	System Architecture	07
2.2	Flow Chart	08
2.3	Use case diagram for Brands(User)	09
2.4	Use case diagram for Admin	09
2.5	Class Diagram	10
2.6	Heatmap	13
3.1	Platform for Execution	16
4.1	RMSE Value	17
4.2	Graphical representation for RMSE value	17
4.3	Final Prediction	18
4.4	Home Page	18
4.5	Admin Login	19
4.6	Brand Signup	19
4.7	Admin Dashboard	20
4.8	Influencers List	20
4.9	Brand(user) Dashboard	21
4.10	Search Influencers	21
4.11	Influencers Information	22
4.12	List (bookmark)	22
4.13	Settings	23
4.14	Dynamic Data	23

Chapter 1

Project Conception and Initiation

1.1 Abstract

Social media networks like Facebook, Twitter, and Instagram have provided a marketing platform for marketers(Brands) to interact with their customers. Through Instagram, Brands can share photos and videos about their products by collaborating with influencers. An influencer is a person who can influence potential buyers of the product or service by promoting it on social media. They make regular posts about that topic on social media and generate massive followings of enthusiastic, engaged people who pay in interest to their views. There are 1 billion active Instagram influencers and brands that need to find the best Influencer for marketing their products will require a tool that will talk about Engagement rates (Engagement rates refer to the number of likes, comments, and shares an influencer gets when posting content online) of the influencers. In this paper, a discussion on how Brands on Instagram can find an Influencer for marketing their product will be done. We will be using machine learning algorithms like Linear Regression, Random Forest and whichever algorithm will give the best result we will use that algorithm for the final prediction.

1.2 Introduction

From the past few decades, the massive use of the internet has led to an increase in online activities such as social media communication, e-commerce, blogging, surfing, etc. This usage has also created a huge impact. On different business Machine learning permits the user to feed an algorithmic program a vast quantity of information and have to analyze and build data driven recommendations and selections supported solely by the computer file. The basic objective of AI (also known as heuristic programming, intelligence or the simulation of psychological feature behavior and this alter computers to perform such intellectual tasks as higher cognitive process drawback solving perception, understanding human communication (in any language, and translation among them). To support this, we are trying to create a system which will be useful for brands which are attempting to collaborate with Instagram influencers based on their engagement rate and growing and declining phases on the premise of their recent activities. Instagram influencers area unit regular Instagram users however with a singular ability to influence others, well established believability, and a considerably giant audience. From a

promoting perspective, an influencer is outlined as an individual with the facility to influence potential patrons of a product or service. Based on that data, internet users in the world have reached 4 billion. This means that more than half the world's population is connected to the internet and 3.196 billion are actively using social media. Indonesia is huge. Social media can be a means of participation and engagement, namely the delivery of information through social media that is directed to encourage public participation and engagement by providing comments, responses, and input to government agencies. The use of social media by the government can provide opportunities for the government to fulfill the basic objectives of democracy, namely transparency, citizen participation, and engagement. Based on the literature review below, many researchers propose measurement methods for social media Facebook while for Instagram it is still small, as was done by Azmi and Budi in 2018 with the research title "Exploring Practice and Engagement of Instagram by Indonesia Government Ministries. The author states that there are two ways in which Instagram users can interact or engage with other users namely likes and comments. We are going to predict engagement rate with the help of machine learning algorithms like Linear Regression, and Random forest. and will see which algorithm gives the best result and will use that algorithm for future predictions. Algorithms will be selected on the basis of lowest rmse values. We will be considering cases with feature selection and tuning hyperparameters. Our end result will be predicting engagement rates of Instagram influencers on that basis brands will decide which influencer it has to collaborate with.

Engagement rate can also be calculated manually with formulas:

$LE = \text{likes} / \text{followers} * 100$

$CE = \text{comment_counts} / \text{followers} * 100$

Total Engagement = $LE + CE$

Where,

LE = Likes Engagement

CE = Comments Engagement

1.3 Objective

- Study Influencer Marketing and its process as one of the emerging fields of marketing in the digital space.
- Overview of Prediction Engagement Rate and its effects and consequences.
- Exploring the practice for the parameters of influence within the application of Prediction to evaluate the effectiveness of the Influencer.
- Explore the opportunities and challenges of using Prediction of engagement rate to evaluate Influencer Marketing.

1.4 Literature review

With the arrival of technology and digital transformation, the Internet has evolved from being a data exchange source from restricted supplies to an open platform wherever on this planet will access and make user generated content. Some user generated content gave rise to Influencers who assist bound selling activities by promoting a product or services and influencing its audience (viewers) to buy/use the product or services (Duncan Brown, 2008). Influencer Marketing is flourishing and wheeling up by giving out effective results and additional selling managers are opting for Influencer Marketing as another to achieve a wider audience on social media with a really short span of your time (Evans, 2017). However, it is necessary for managers to pick the correct influencers for their merchandise and services to broadcast the correct image within the minds of the audience.

In [1], Sentiment Analysis and classification of Restaurant Reviews using Machine learning.

This model is to analyze the customer reviews about various restaurants across Karachi. The contribution of this research is twofold. First, it performs sentiment analysis and classifies each comment as positive, negative.

In [2], Measurement of Engagement Rate in Instagram.

The author took an explorative engagement approach dealing with government data processing. Social networking(online); Internet; learning (artificial intelligence); mobile computing; computer- aided instruction; smart cities; Big Data; educational institutions; Internet of Things.

In [3], Measuring Instagram activity and Engagement Rate of Hospital A comparison Before and During COVID-19 pandemic.

This model identifies the hospitals' frequency of posts in their Instagram account and the engagement rate before and during the COVID-19 pandemic.

In [4], Sentiment/opinion Review analysis Detecting spams from the good ones!
We go through this in a step by step format of different papers and summarize for other readers how we can identify the correct emotions and differentiate between the real and fake reviews.

In [5], An Effective Machine learning Approach for Sentiment Analysis on Popular Restaurant Reviews in Bangladesh.

We built a model using natural language processing techniques and machine learning algorithms to automate the approach of classifying reviews on around 200 popular restaurants of Bangladesh as Satisfactory or Poor.

In [6], Exploring practices and engagement of Instagram by Indonesian Government Ministries.

The results of analysis of ministry Instagram data.

In [7], Social Network Influencers Engagement Rate Algorithm Using Instagram Data.

Research has been done about related case studies and existing algorithms to get a better understanding of how to use available metrics such as likes, comments, and followers, to calculate the engagement rate of Influencers in the most accurate way.

Published Year	Author	Title	Methodology used	Algorithm used
2020	Kanwal Zahoor; Narmeen Zakaria Bawany; Soomaiya Hamid	Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning	The focus of this paper is to analyze the customer reviews about various restaurants across Karachi. The contribution of this research is twofold. First, it performs sentiment analysis and classifies each comment as positive, negative.	Random forest algorithm (95% accuracy)
2019	Arry Akhmad Arman	Measurement of Engagement Rate in Instagram	The main objective in this study is to take an explorative engagement approach dealt with: government data processing; social networking (online); Internet; learning (artificial intelligence); mobile computing; computer aided instruction; smart cities; Big Data; educational institutions; Internet of Things	Random forest algorithm (95% accuracy)
2020	Badra Al Aafa; Wahyu Sulistiadi; Faizah Abdullah Djawas	Measuring Instagram Activity and Engagement Rate of Hospital: A Comparison Before and During COVID-19 Pandemic	The study aimed to identify the hospitals' frequency post in their Instagram account and the engagement rate before and during the COVID-19 pandemic.	Mann whitney test
2019	Yshika Agarwal; Dilip Kumar Sharma; Rahul Katarya	Sentiment/Opinion Review Analysis: Detecting Spams from the good ones!	In this article, we go through this in a step by step format of different papers and summarize for other readers how we can identify the correct emotions and differentiate between the real and fake reviews.	coupled stochastic mode
2019	S M Asiful Huda; Md Mohiuddin Shoikot; Md Anower Hossain; Ishrat Jahan Ila	An Effective Machine Learning Approach for Sentiment Analysis on Popular Restaurant Reviews in Bangladesh	In this study, we built a model using natural language processing techniques and machine learning algorithms to automate the approach of classifying a review on around 200 popular restaurants of Bangladesh as Satisfactory or Poor.	Support Vector Machine
2018	Achmad Fauzi Azmi; Indra Budi	Exploring practices and engagement of Instagram by Indonesia Government Ministries	The results of analysis of ministry Instagram data revealed that the engagement level of the people does not correlate with the frequency of Instagram ministries post.	Linear regression
2018	Roy Ling Hang Yew; Syamimi Binti Suhaidi; Pristee Seewoochurn; Venantius Kumar Sevamalai	Social Network Influencers' Engagement Rate Algorithm Using Instagram Data	Research has been done about related case studies and existing algorithms to get a better understanding of how to use available metrics such as likes, comments and followers, to calculate the engagement rate of influencers in the most accurate way.	Xg-Boost

Figure 1.1: Literature Review

1.5 Problem Definition

While everything seems to be working out in numbers there still seems to be multiple issues Brands and Influencers are facing. Some of the issues are as follows: time management/Bandwidth, finding influencers, processing payments to influencers, managing contracts, etc. Out of these, finding influencers is one of the difficulties which the brand faces the most. and even though the campaign is happening on the platform the Payment and tracking of the campaign is happening through several other companies like Influent, scrunch, etc. This is one of the gaps that can be filled to make things a little simple for Business users. Brands on Instagram need to collaborate with Instagram influencers to promote their product. So, the challenge for the brand is to find the correct influencer with the highest engagement rate. Machine learning can be used for making a brand's work easier.

1.6 Scope

1. The proposed system will predict the engagement rates correctly on the basis of data input like likes, comments, shares, posts, etc.
2. This application will be used to calculate engagement rates of Instagram Influencers. In this project, we have to help brands find the best instagram user for branding its products.

This platform makes brands work easier by calculating engagement rates.

3. The reason for developing such a system is, there are millions of users on instagram but out of them, which user is constantly active, which user has more shares, and more response by audience, that's the question. So to help brands find the users having the highest engagement rate will help them for better sales of their product. Here, on the basis of engagement rate we define whether the user will be in the growing/ declining phase.
4. If a user in a growing phase means that the audience interaction with that user will continue like this. If in a declining phase means that the audience interaction with that user will decrease.

1.7 Technology Stack

Front- End

- **HTML** - HTML (Hypertext Markup Language) is the code that is used to structure a web page and its content. HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets and scripting languages such as JavaScript.
- **CSS** - "Cascading Style Sheets" is a code that is used to layout and structure a web page. It is used for describing the presentation of a document written in a markup language such as HTML. There are three types of CSS:
 1. Inline CSS.
 2. Internal or Embedded CSS.
 3. External CSS.
- **JavaScript** - JavaScript is the world's most popular programming language. JavaScript is the programming language of the Web. JavaScript is easy to learn. JavaScript, often abbreviated JS, is a programming language that is one of the core technologies of the World Wide Web, alongside HTML and CSS.
- **Python** is a general purpose interpreted, interactive, object oriented, and high level programming language. Python is designed to be highly readable. Python has become one of the most popular programming languages in the world in recent years. It's used in everything from machine learning to building websites and software testing. It can be used by developers and non developers alike. Python can build a wide range of different data visualizations, like line and bar graphs, pie charts, histograms, and 3D plots.
- **Scikit-learn** is a free software machine learning library for the Python programming language. Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. Scikit-learn is an indispensable part of the Python machine learning toolkit at JPMorgan. It is very widely used across all parts of the bank for classification, predictive analytics, and very many other machine learning tasks. Its straightforward API, its breadth of algorithms, and the quality of its documentation combine to make scikit-learn simultaneously very approachable and very

powerful.

- **NumPy** stands for Numerical Python. NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, Fourier transform, and matrices. Using NumPy, a developer can perform the following operations

Mathematical and logical operations on arrays. Fourier transforms and routines for shape manipulation. Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

- **Matplotlib** is a low level graph plotting library. Matplotlib is a cross platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications.
- **Numeric**, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray was also developed, having some additional functionalities. In 2005, Travis Oliphant created the NumPy package by incorporating the features of Numarray into the Numeric package.

Back-End

- **PhpMyadmin** - phpMyAdmin is a free and open source administration tool for MySQL and MariaDB. As a portable web application written primarily in PHP, it has become one of the most popular MySQL administration tools, especially for web hosting services.

1.8 Benefits for Environment and Society

- You can develop a more insightful, data based marketing strategy. Nothing beats data based strategy.
- Understand your customers.
- Measure your marketing campaign.
- Take a look at brand perception.
- Find industry leaders and influencers.
- Give an extra boost to your customer service.

Chapter 2 Project Design

2.1 Proposed System

- The presentation layer which is concerned with delivering the data to the user. So, those area units typically the UI connected issues that area units addressed at the presentation layer. Here we tend to typically have hypertext mark-up language, CSS, and JavaScript.
- The business logic layer on the opposite hand is bothered a lot about the information, the information validation, the dynamic content process, and generating the content to be delivered to the user. Here we tend to typically have Node, Python,c++, or PHP.
- This is insured behind the scenes with the information persistence layer or the information access layer. So, this is often involved with however we tend to store and move with the information, generally within the type of information, and access this informa- tion through an API. Therefore, the business logic layer interacts with the information like MongoDB or PostgreSQL.

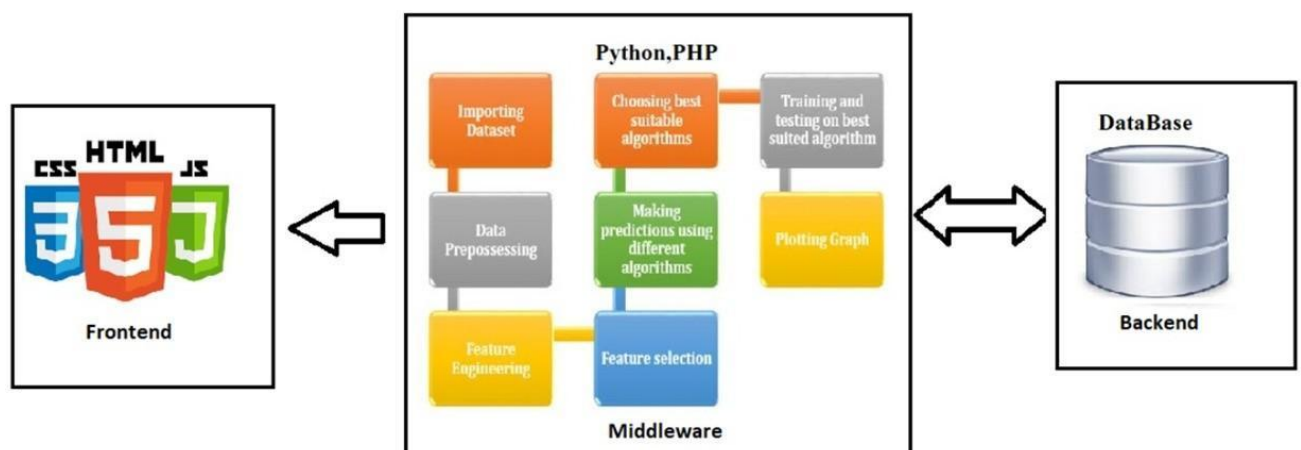


Figure 2.1: System Architecture

2.2 Design(Flow of Modules)



Figure 2.2: FLOW Chart

- Importing Dataset
- Data Preprocessing
- Feature Engineering
- Choosing best suitable algorithm
- Making prediction using different algorithms
- Feature Selection
- Training and testing on best suited algorithm
- Plotting graph

Above steps are explain in Module 1-4

2.3 Description Of Use Case

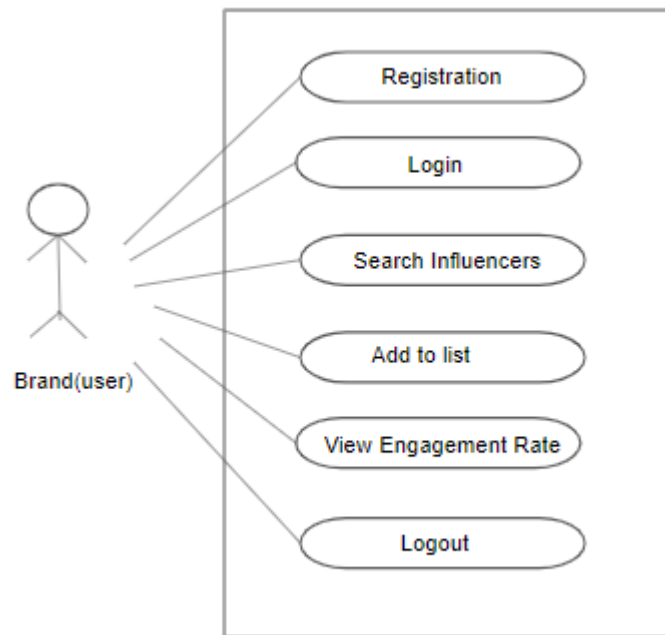


Figure 2.3: Use case Diagram for Brand(user)

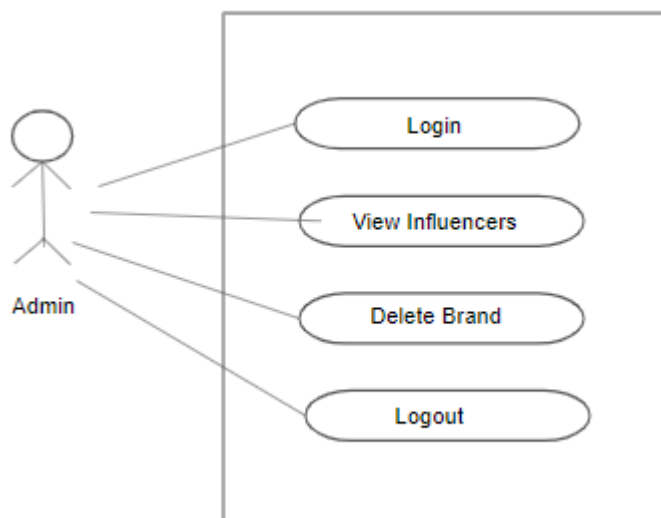


Figure 2.4: Use case Diagram for Admin

2.4 Class Diagram

Admin can add new Brands, Influencer and also remove Brand, Influencer. Admin can also add new admin.

Home admin can login, brand can login and signup.

Users can search the influencer, engagement rate, change password, and view information.

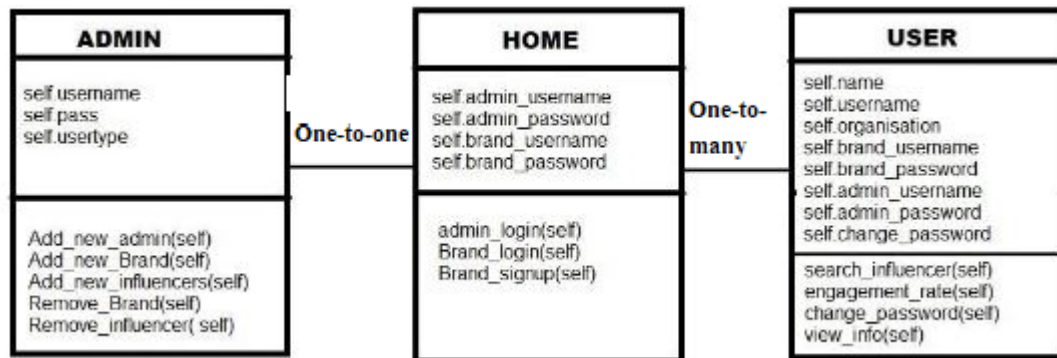


Figure 2.5: Class Diagram

2.5 Modules

2.5.1 Module 1

Importing Dataset

- We have taken the dataset from Github. It has 11 columns and 52687.

Data Preprocessing

- Data Preprocessing is that step in which the data gets transformed, or Encoded, to bring it to such a state that now the machine can easily parse it.
- Preprocessing data starts from data cleansing, feature engineering, feature selection etc until the data is ready to be consumed by the Machine Learning model.
- Four stages: data cleaning, data integration, data reduction, and data transformation.

2.5.2 Module 2

Feature Engineering

- Feature engineering is the process of selecting, manipulating, and transforming raw data into features that can be used in supervised learning.

Feature selection

- The next step is Feature Selection. The method I use is to look at the Correlation Coefficient between the Predictor and the Target Features.
- The goal of feature selection in machine learning is to find the best set of features that allows one to build useful models of studied phenomena.
- There are supervised and unsupervised techniques.
- From a taxonomic point of view, these techniques are classified as under:
 - A. Filter methods
 - B. Wrapper methods
 - C. Embedded methods
 - D. Hybrid methods

Correlation Coefficient

- Correlation is a measure of the linear relationship of two or more variables.
- If two variables are correlated, we can predict one from the other. Thus, if two features are correlated, the model only really needs one of them, as the second one does not add additional information.

2.5.3 Module 3

Making predictions using different algorithms

- Modeling uses Machine Learning Algorithm (Linear Regression, Random Forest) and also does some Tuning Hyperparameters

Hyperparameters Tuning

- A Machine Learning model is defined as a mathematical model with a number of parameters that need to be learned from the data. By training a model with existing data, we are able to fit the model parameters.

GridSearchCV

- In the GridSearchCV approach, the machine learning model is evaluated for a range of hyperparameter values. This approach is called GridSearchCV, because it searches for the best set of hyperparameters from a grid of hyperparameters values.

In our project we will be considering 4 cases:

- Modeling without Feature Selection and without Tuning Hyperparameters.
- Modeling without Feature Selection and with Tuning Hyperparameters.
- Modeling with Feature Selection and without Tuning Hyperparameters.

- Modeling with Feature Selection and with Tuning Hyperparameters.

Metrics evaluation that we will be using is Root Mean Squared Error (RMSE)

- Root mean square error or root mean square deviation is one of the most commonly used measures . It shows how far predictions fall from measured true values using Euclidean distance. RMSE is commonly used in supervised learning applications, Lower values of RMSE indicate better fit.

Choosing best suitable algorithms

- The algorithm which will give lowest rmse value will be chosen for final prediction and results will be evaluated.

2.5.4 Module 4

Training and testing on best suited algorithm

- Whichever RMSE value is least that will be better fitted so we will choose that algorithm to make our final predictions. Our main target is to predict the engagement rates of different users and based on that the user will be categorized in to "Growing" or "Declining" phase. Brands will be able to select Influencers based on which phase they are.

Plotting Graph

A heatmap contains values representing various shades of the same color for each value to be plotted. Usually the darker shades of the chart represent higher values than the lighter shade.

```
feature = ['username', 'year_month', 'n_post', 'n_post_01', 'n_img_post', 'avg_likes',
'avg_likes_01','avg_likes_02','avg_likes_03','sum_capt_wo_punct','sum_n_words',
'sum_avg_char_words','sum_likes_engagement','sum_total_engagement',
'avg_likes_engagement','avg_total_engagement','avg_total_engagement_01',
'avg_total_engagement_02', 'avg_total_engagement_03', 'y_month_01']
```

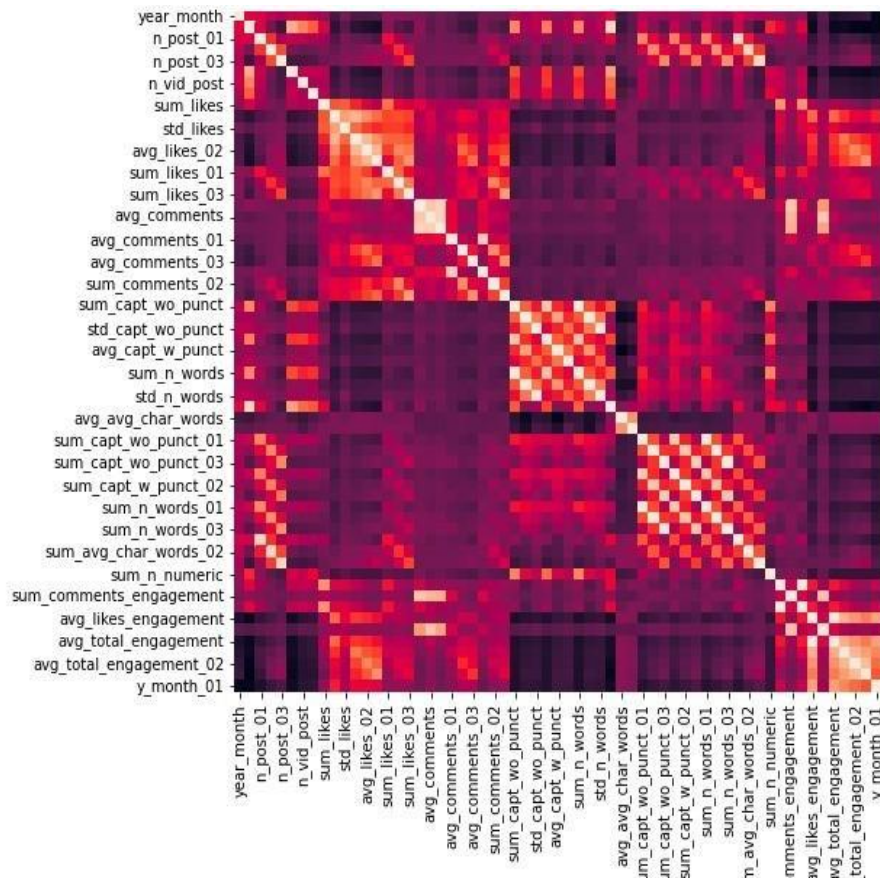


Figure 2.6: Heatmap

Chapter 3 Implementation

3.1 Proposed system-implementation

1. First brand can login to the website,
2. Then user dashboard appear in that brand details is shown
3. After the user dashboard there is an search influencer
4. In search influencer we can search any influencer with their category
5. We can also bookmark the influencers that will be stored in the list tab.
6. In the settings tab we can change a password for the brand. After that we logout from the brand dashboard.
7. After that admin dashboard there is a Brand. Where each Brand is shown
8. After brand there is an Influencer tab. In influencer tab we can see the list of influencer
9. In the settings tab we can change a password for the Admin. After that we logout from the brAdmin dashboard.

3.1.1 Policies

The First task in our project was to search the topic. We need to search for a project which is practical oriented and theoretically aligned. After reading all the descriptions and discussing it with others a project topic is finalized.

After finalizing with others, our next task is to identify the problem. Problem identification involves a lot of background work in the general area of the problem. Normally it calls for the use of prior experience, typically experience we may not yet have. For identifying the task, we need to read as much base paper as we can. After identifying the problem we are now able to design our system.

After the Designing is finalized we now start with the actual coding part. We will split the coding into small modules. When each module is fully qualified, we will integrate into one system. After the implementation of our idea, we must test for the bugs by using all possible test scenarios and test cases. After the testing phase is done, we are now ready to deploy our project.

3.1.2 Algorithms

- **Linear Regression:**

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Linear regression analysis is used to predict the value of one variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. In Linear Regression, we accept that there is a connection between an autonomous variable vector and the needy target variable. By utilizing the free parameters, we can anticipate the objective variable.

- **Random Forest :**

Random forest is a supervised learning algorithm. Random forest algorithms can be used for both classifications and regression tasks. Random forest is a flexible, easy-to-use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It provides higher accuracy through cross validation. Random Forest forms several decision trees and integrates them so that they can predict more accurately. When the tree starts spreading it creates more randomness in the model. Due to the wide diversity of Random Forest provides better results and prevents overfitting. It takes less training time as compared to other algorithms. It can also maintain accuracy when a large proportion of data is missing.

3.1.3 Pseudo code

Login Authentication

```
1.s = Session()
2.query = s.query(User).filter(User.username.in ([POST USERNAME]),
3.User.password.in ([POST PASSWORD]))
4.result = query.first()
5.if result:
6.session['logged in'] = True
7.else:
8.flash('wrong password!')
9.return index()
```

```

10.logout.php:
11.session['logged in'] = False
12.return index()

```

Engagement Rate Prediction

```

1.final_predictions=best_model.predict(data_predict)
2.df_prediction_2['avg_engagemnt_prediction_july'] = final_prediction
3.df_prediction_3['category'] = df_prediction_3['difference_june_july'].apply(lambda x:
'Growing' if x > 0 else 'Declining')
4.df_prediction_3.to_csv('final_result.csv', index=False)

```

3.1.4 Platforms for execution

Sr. No	Software Type	Software Name
1.	Anaconda Navigator	Jupyter Notebook, Visual Studio Code
2.	Database	MySQL (phpMyadmin)
3.	Web Browser	Google Chrome, Chrome Driver

Figure 3.1: Platform For Execution

Chapter 4 Result

RMSE Values - Testing Data				
Alogorithms	Modelling without feature selection without Tuning	Modelling without feature selection with Tuning	Modelling with feature selection without Tuning	Modelling with feature selection with Tuning
Linear Regression	2.02214	2.0024	1.9994	1.9819
Random Forest	1.9079	1.68934	1.9563	1.5016 (MINIMUM)

Figure 4.1:RMSE Value

In Modeling without feature selection and without tuning: RMSE value of Linear regression is greater than Random forest In Modeling without feature selection with tuning: RMSE value of Linear regression is greater than Random forest In Modeling with feature selection without tuning: RMSE value of Linear regression is almost equal to Random forest In Modeling with feature selection with tuning: RMSE value of Linear regression is greater than Random forest among all the above RMSE value Random forest give the least value that's why we select that algorithm for final prediction.

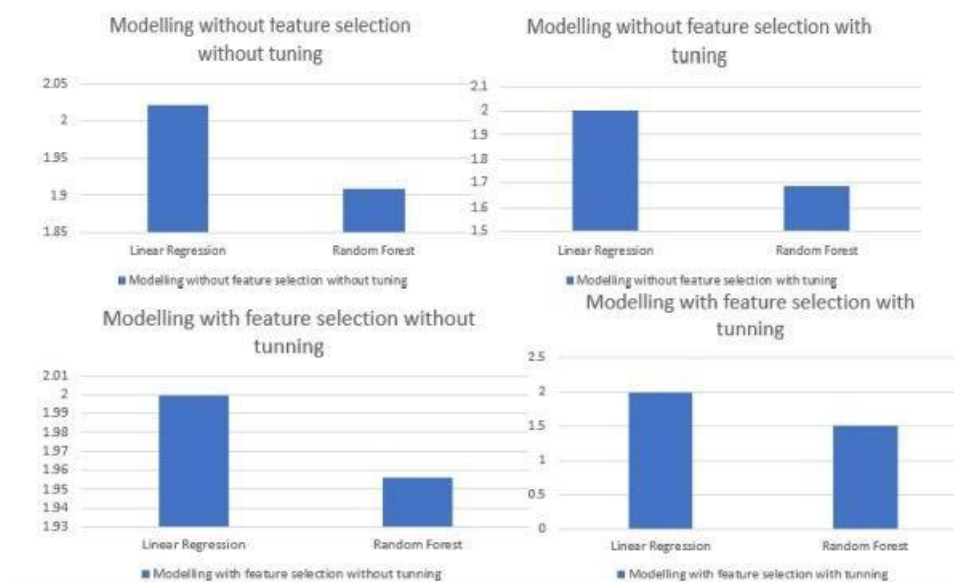


Figure 4.2: Graphical representation of RMSC value


```
df_prediction_3.sample(10)
```

	username	avg_total_engagement_june	prediction_avg_total_engagement_july	difference_june_july	category
45	angelagilsha	3.06	3.11	0.05	Growing
8	adiba.knza	16.19	14.41	-1.78	Declining
605	vidialdiano	4.71	4.60	-0.11	Declining
29	alicenorin	1.55	1.90	0.35	Growing
240	geoff_max	1.27	1.35	0.08	Growing
459	pathdaily_story	0.64	0.95	0.31	Growing
538	sherlyannavita	1.69	1.76	0.07	Growing
456	pammybowie	2.63	2.51	-0.12	Declining
367	madkucil	4.76	4.41	-0.35	Declining
34	altynbekova_20	9.01	8.44	-0.57	Declining

Figure 4.3: Final Prediction

This figure shows the output of the predicted value of next month's engagement rate. And if the difference between previous month's and predicted next month's engagement rate is positive, it means it is a growing phase, else a declining phase.

Front End Result:

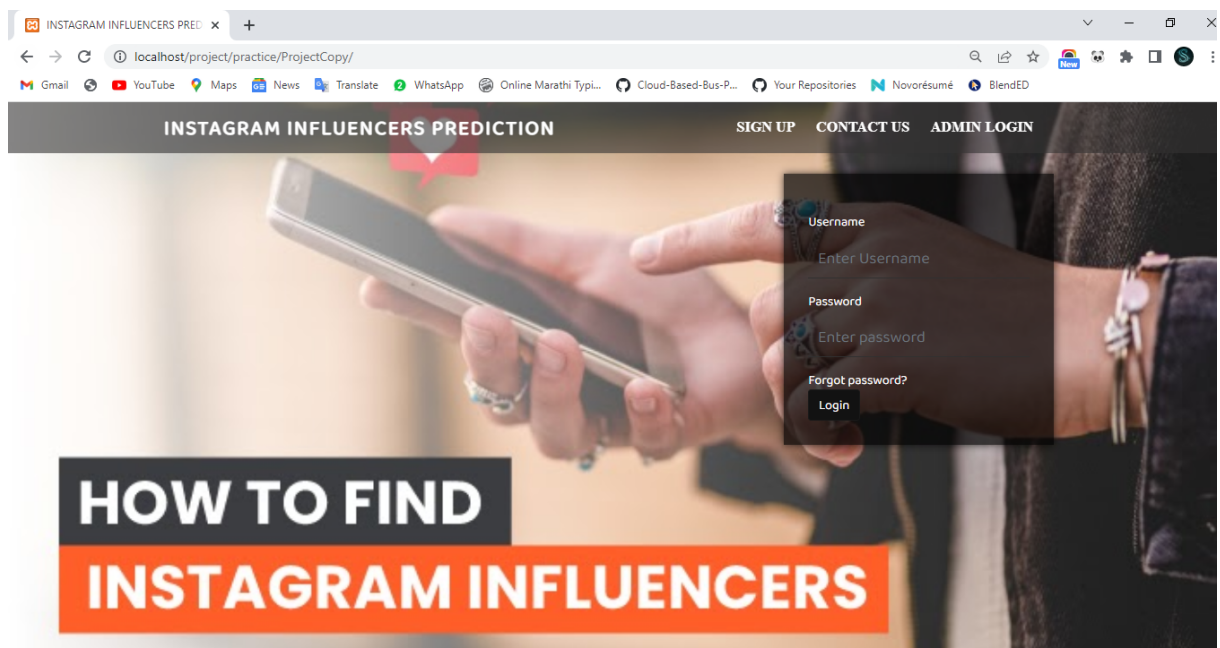


Figure 4.4: Home Page

This is the main Home Page of our project, where the Brand(user) can login and search influencers. Also there is option for Admin Login and Brand Signup.

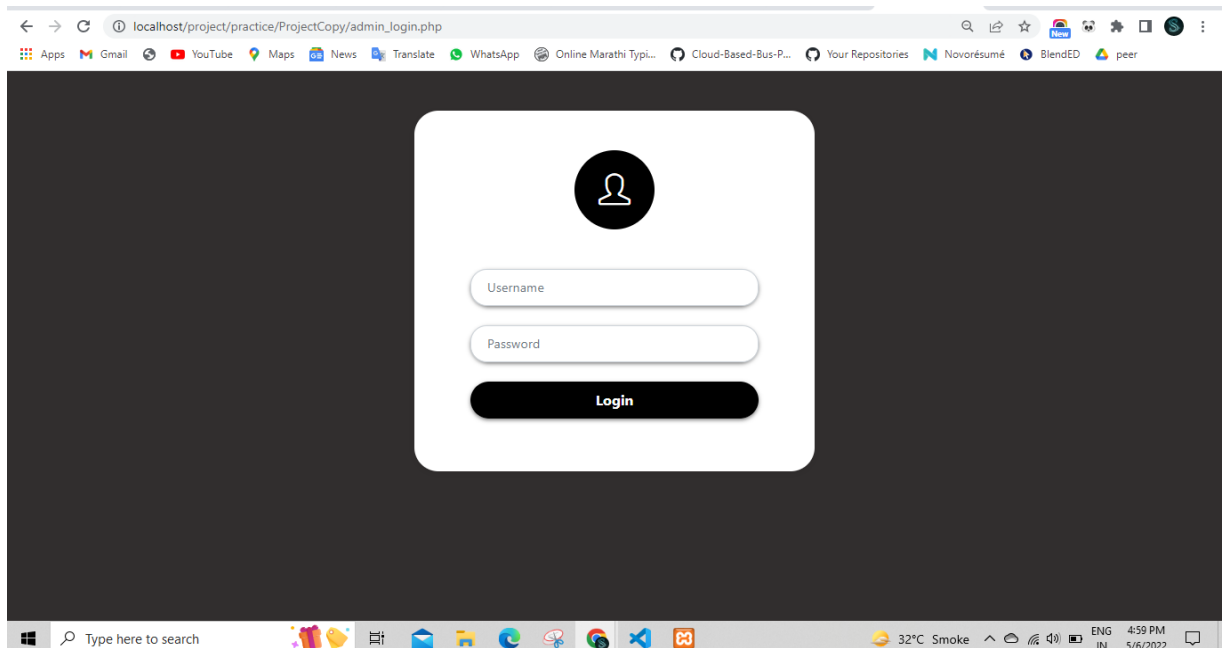


Figure 4.5:Admin Login

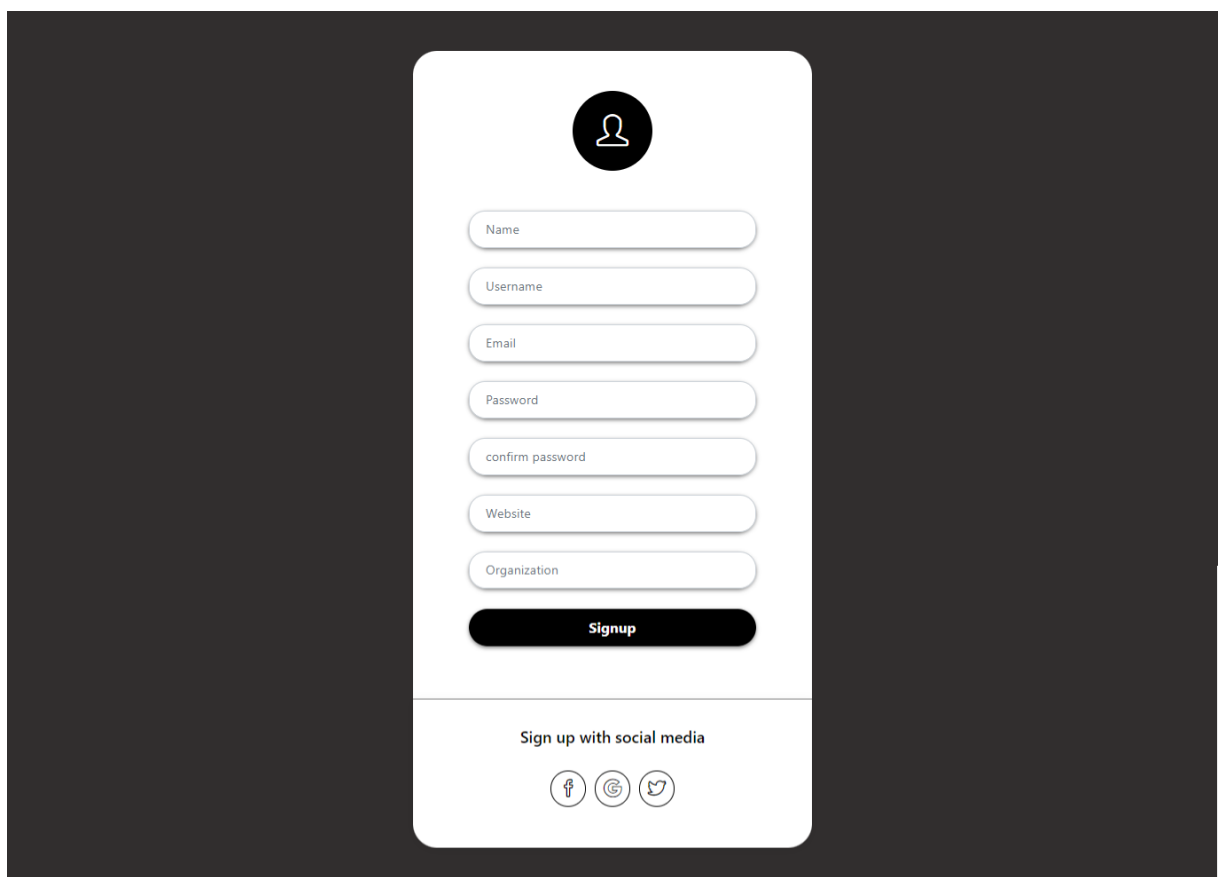


Figure 4.6:Brand Signup

The above images are of Admin Login and Brand Signup

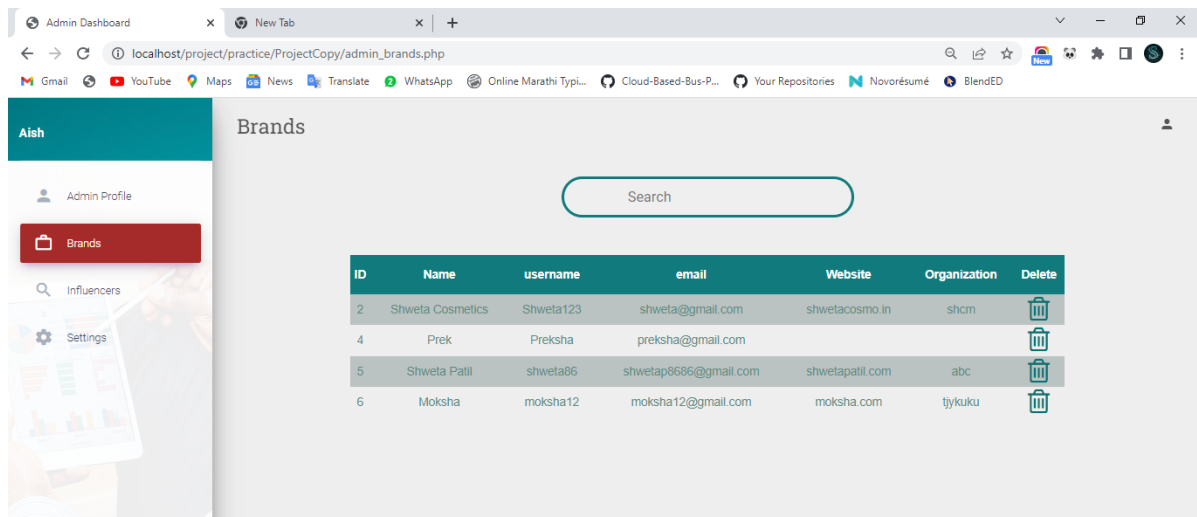


Figure 4.7:Admin Side - List of Brands

This tab shows all the list of brands which have signup and created account. Admin can view details of brands and can also delete brands.

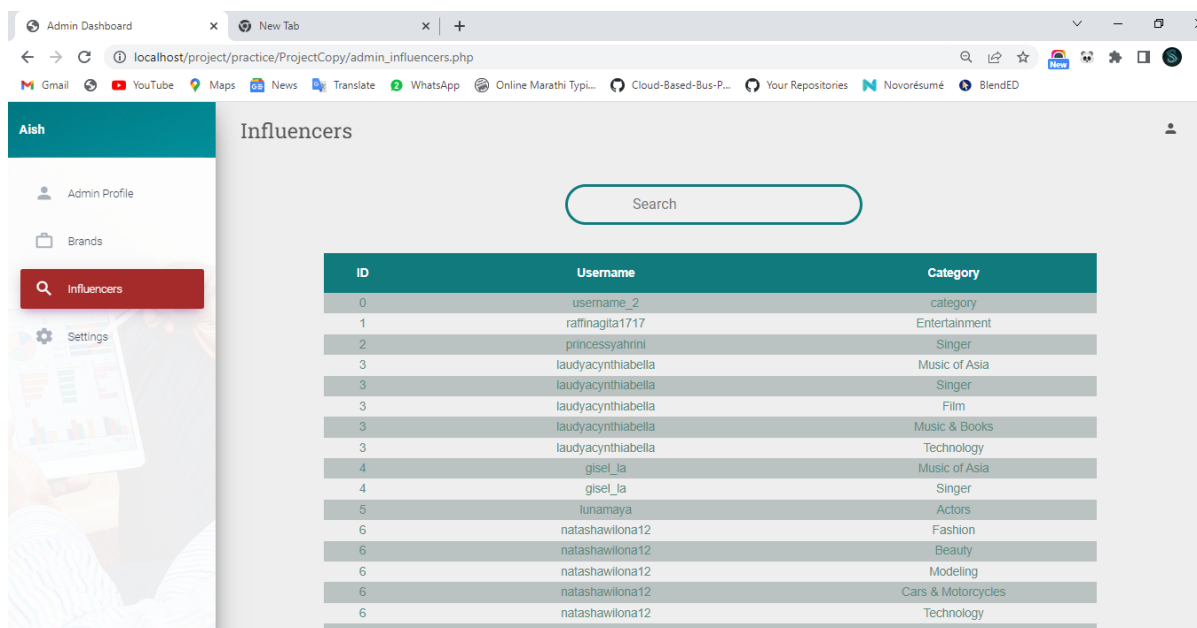


Figure 4.8: Influencers List

This tab shows all the list of influencers present in the database along with their category.

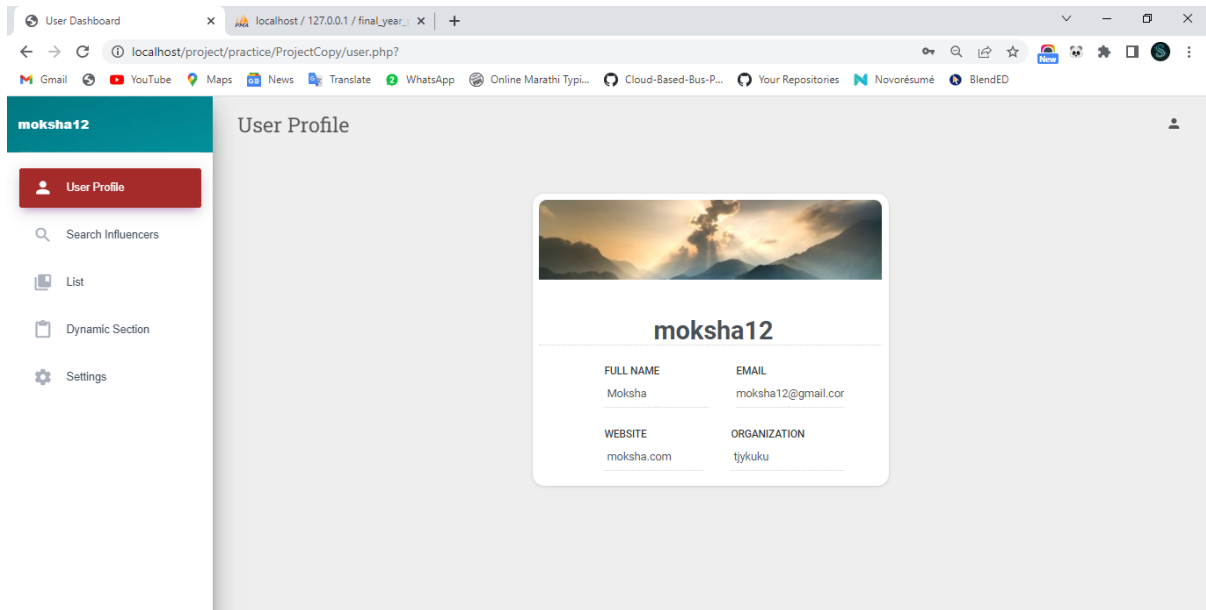


Figure 4.9:Brands(user) Dashboard

This is user profile page where user can see all it's details filled while signup.

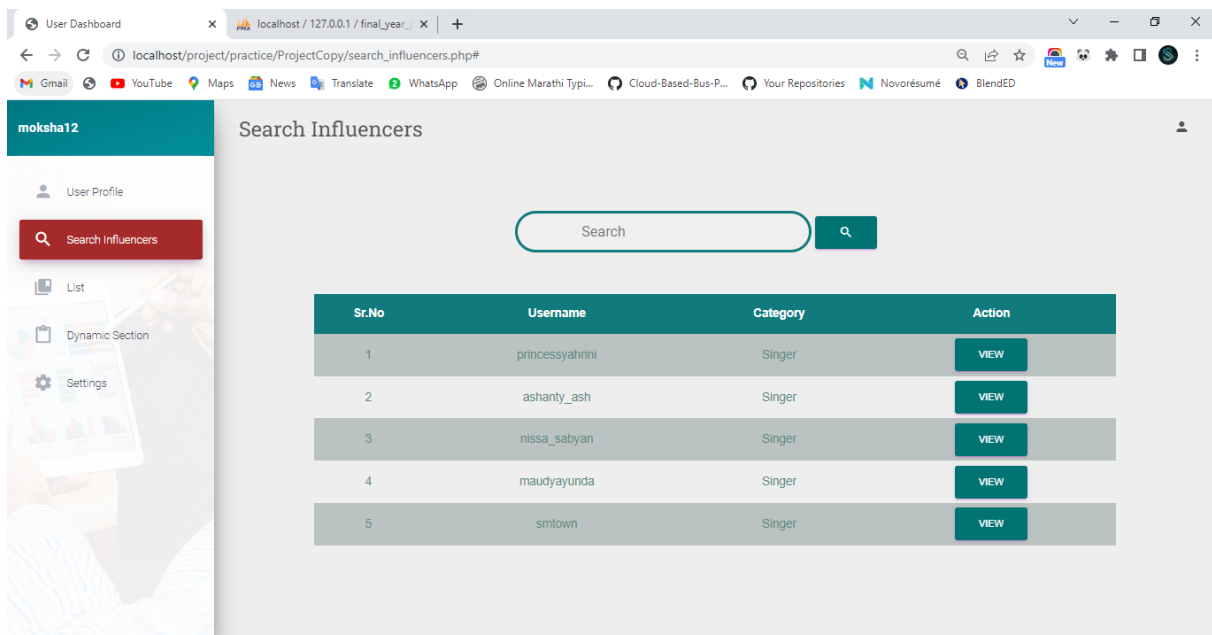


Figure 4.10:Search Influencers

This search influencers page at Brand(user) side shows the details of the influencers which can be searched according to category. The Brand(user) can also view the details of each influencers.

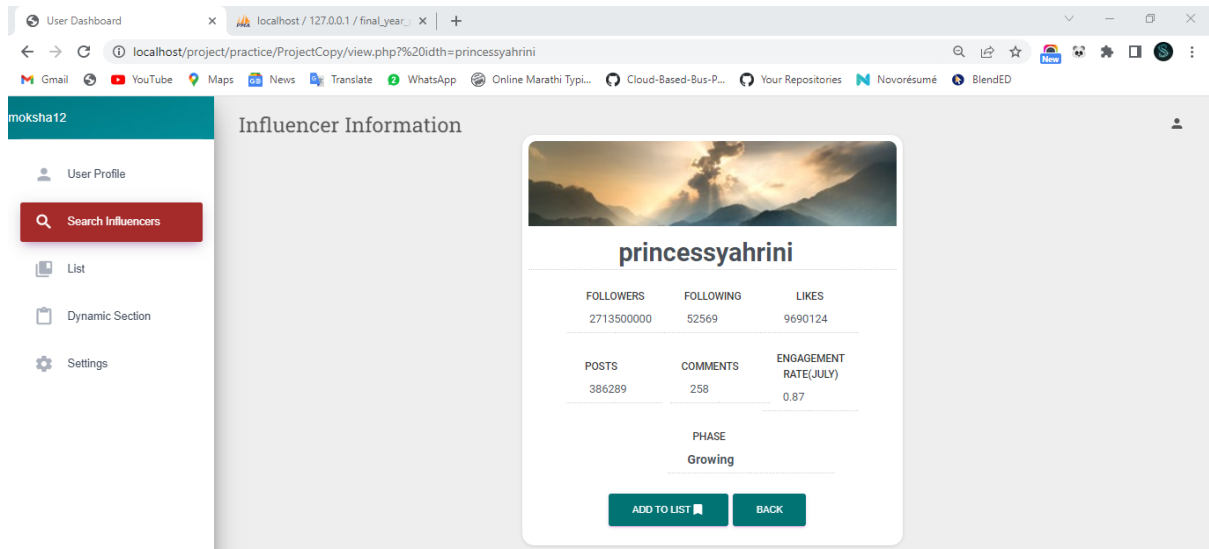


Figure 4.11:Influencers Information

On clicking the view button on search influencers page, you get on this page where you can view all the details of influencer and if Engagement Rate is growing then brand can also add it to your favorite list.

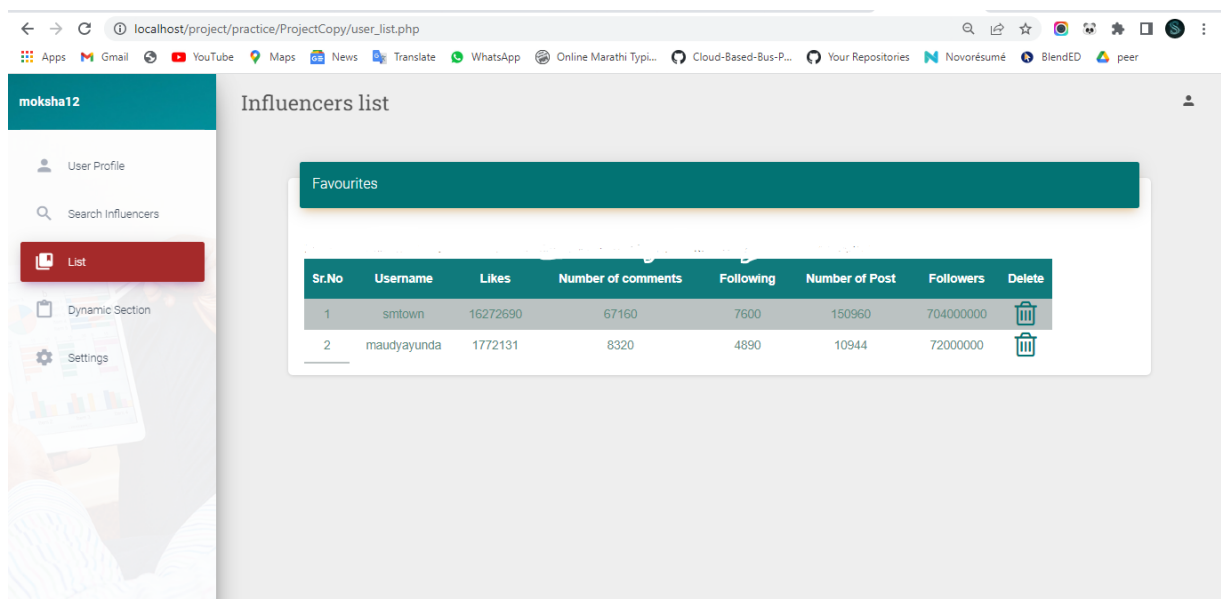


Figure 4.12:List (bookmark)

On adding to favorite list , we can view in this tab, Brand(user) can also delete the influencer.

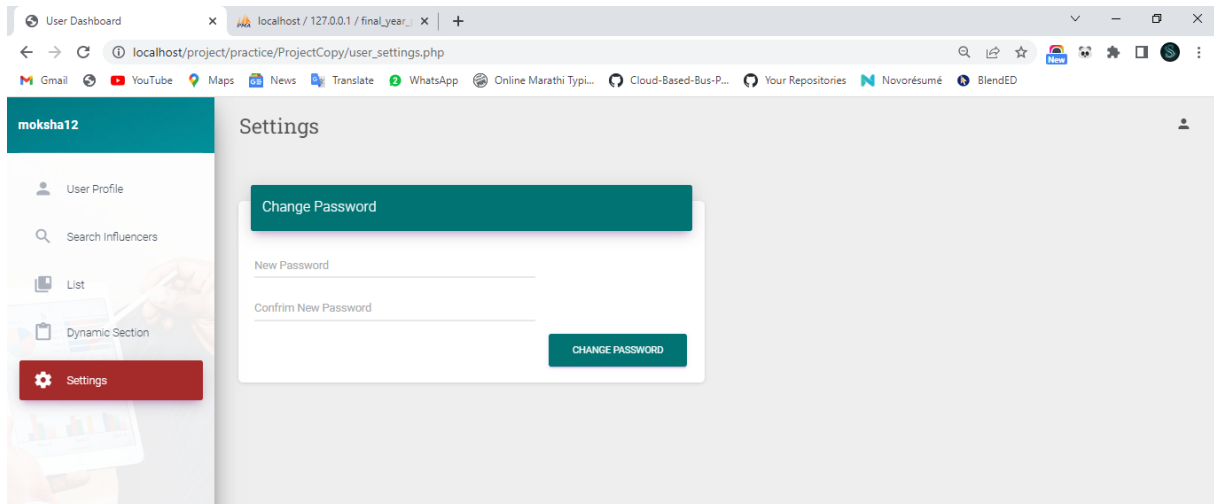


Figure 4.13:Settings

Brands(user) can change/update the password of the account using settings tab.

Dynamic Data:

```
C:\Users\moksha shah\Desktop>python test2.py
Enter username:aishwarya_7818

DevTools listening on ws://127.0.0.1:56327/devtools/browser/8d84c014-a2a2-4101-9d66-ff53f65e5074
9 226 151
{'Followers': '226', 'Following': '151', 'Posts': '9'}
[11088:19148:0428/212523.901:ERROR:device_event_log_impl.cc(214)] [21:25:23.902] USB: usb_device_handle_win.cc:1049 F
ailed to read descriptor from node connection: A device attached to the system is not functioning. (0x1F)
[11088:19148:0428/212523.903:ERROR:device_event_log_impl.cc(214)] [21:25:23.904] USB: usb_device_handle_win.cc:1049 F
ailed to read descriptor from node connection: A device attached to the system is not functioning. (0x1F)
[11088:19148:0428/212523.933:ERROR:device_event_log_impl.cc(214)] [21:25:23.933] Bluetooth: bluetooth_adapter_winrt.c
c:1165 RequestRadioAccessAsync failed: RadioAccessStatus::DeniedByUserWill not be able to change radio power.

DevTools listening on ws://127.0.0.1:56386/devtools/browser/d0b803ea-94d0-4768-bc56-00c01359f23b
[18244:2280:0428/212538.317:ERROR:device_event_log_impl.cc(214)] [21:25:38.317] USB: usb_device_handle_win.cc:1049 Fa
iled to read descriptor from node connection: A device attached to the system is not functioning. (0x1F)
[18244:2280:0428/212538.317:ERROR:device_event_log_impl.cc(214)] [21:25:38.318] USB: usb_device_handle_win.cc:1049 Fa
iled to read descriptor from node connection: A device attached to the system is not functioning. (0x1F)
[18244:2280:0428/212538.337:ERROR:device_event_log_impl.cc(214)] [21:25:38.337] Bluetooth: bluetooth_adapter_winrt.c
c:1165 RequestRadioAccessAsync failed: RadioAccessStatus::DeniedByUserWill not be able to change radio power.
['https://www.instagram.com/p/CAE9tV9DvE4/', 'https://www.instagram.com/p/B5pG0gal4FN/', 'https://www.instagram.com/p/
B2SAImhnpaO/', 'https://www.instagram.com/p/B2RN6ITHxX6/', 'https://www.instagram.com/p/BzqW5Nkn85S/', 'https://www.
instagram.com/p/Bsg3c3xID0B/', 'https://www.instagram.com/p/BsdccxQfR0a/', 'https://www.instagram.com/p/BsdbIusfj-G/
', 'https://www.instagram.com/p/BsdXc45FBmP/']
[11088:12908:0428/212602.341:ERROR:util.cc(126)] Can't create base directory: C:\Program Files\Google\GoogleUpdater
[18244:19372:0428/212629.749:ERROR:util.cc(126)] Can't create base directory: C:\Program Files\Google\GoogleUpdater
['63', '66', '79', '60', '74', '48', '50', '49', '60']
Total Likes 549
```

Figure 4.14:Dynamic Data

This is the dynamic data which we have fetched by entering the username. With the help of selenium web browser, you can create an automated application. On running the code, you can enter insta handle. Followers, following, no of posts, post links, comment-counts and total number of likes will be fetched. This data can be feeded to the model for prediction of engagement rate.

Chapter 5

Conclusions and Future Scope

5.1 Conclusion

We have looked at a dataset containing information of Instagram Influencers. Social media networks like Facebook, Twitter, and Instagram have provided a marketing platform for marketers (Brands) to interact with their customers. To help brands in finding the best influencer suitable for their product is the major goal of our project. And this goal has been achieved with better results using various Machine Learning Algorithms. This will help the Brands to find influencers with maximum engagement rates and they can decide whether to collaborate with them or not. Among all the above algorithms used, random forest is the best suitable algorithm for predicting Instagram Influencers.

5.2 Future Scope

- This project can be easily implemented for static datasets containing information about Instagram Influencers.
- It can be further extended to work on a dynamic dataset in which the real time data will be fetched and after modeling, results will be shown to the user in the form of API.
- A recommendation system can be added, which can recommend brands about the further influencers of similar categories based on the brand's interest.

Chapter 6

References

- [1] Zahoor Kanwal, Narmeen Zakaria Bawany and Soomaiya Hamid. "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning.",2020 21st International Arab Conference on Information Technology(ACIT).IEEE 10.1109/ACIT50332.2020.9300098.
- [2] Arry Akhmad Arman; Agus Pahrul Sidik, "Measurement of Engagement Rate in Instagram" (Case Study: Instagram Indonesian Government Ministry and Institutions), 2019 International Conference on ICT for Smart Society (ICISS), DOI: 10.1109/ICISS48059.2019.8969826.
- [3] Badra Al Aufa; Wahyu Sulistiadi; Faizah Abdullah Djawas, "Measuring Instagram Activity and Engagement Rate of Hospital: A Comparison Before and During COVID- 19 Pandemic", 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI),DOI: 10.1109/ISRITI51436.2020.9315490.
- [4] Yashika Agarwal; Dilip Kumar Sharma; Rahul Katarya, "Sentiment/Opinion Review Analysis: Detecting Spams from the good ones!", 2019 4th International Conference on Information Systems and Computer Networks (ISCON),DOI: 10.1109/IS- CON47742.2019.9036249.
- [5] S M Asiful Huda; Md Mohiuddin Shoikot; Md Anower Hossain; Ishrat Jahan Ila, "An Effective Machine Learning Approach for Sentiment Analysis on Popular Restaurant Reviews in Bangladesh", 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS).
- [6] Achmad Fauzi Azmi; Indra Budi, ""Exploring practices and engagement of Instagram 10 by Indonesia Government Ministries", 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE).
- [7] Roy Ling Hang Yew; Syamimi Binti Suhaidi; Prishtee Seewoochurn; Venantius Kumar Sivamalai, "Social Network Influencers Engagement Rate Algorithm Using Instagram Data" ,2018 Fourth International Conference on Advances in ComputingCommunicationAutomation(ICACCA).

Chapter 7

Bibliography

<https://fbohz.com/posts/2020-08-10-threetierfullstack>

<https://stackoverflow.com>

<https://www.python.org>

<https://www.edureka.co/blog/how-to-run-selenium-in-chrome>

<https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86>

<https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/#:~:text=XGBoost%20is%20an%20algorithm%20that,designed%20for%20speed%20and%20performance.>

<https://grin.co/engagement-rate-calculator>

<https://phlanx.com/engagement-calculator>

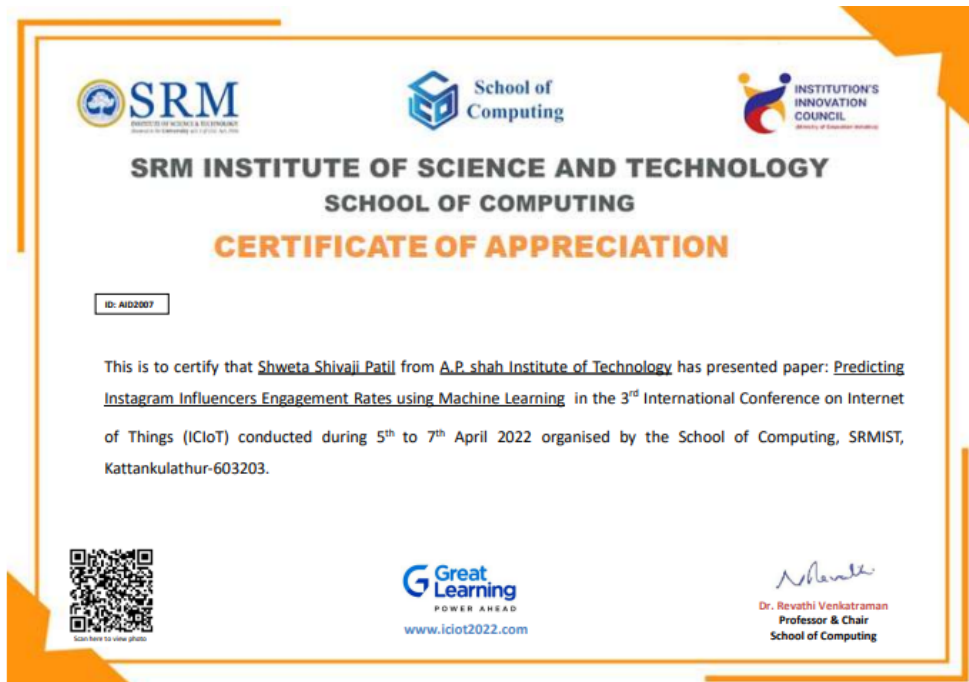
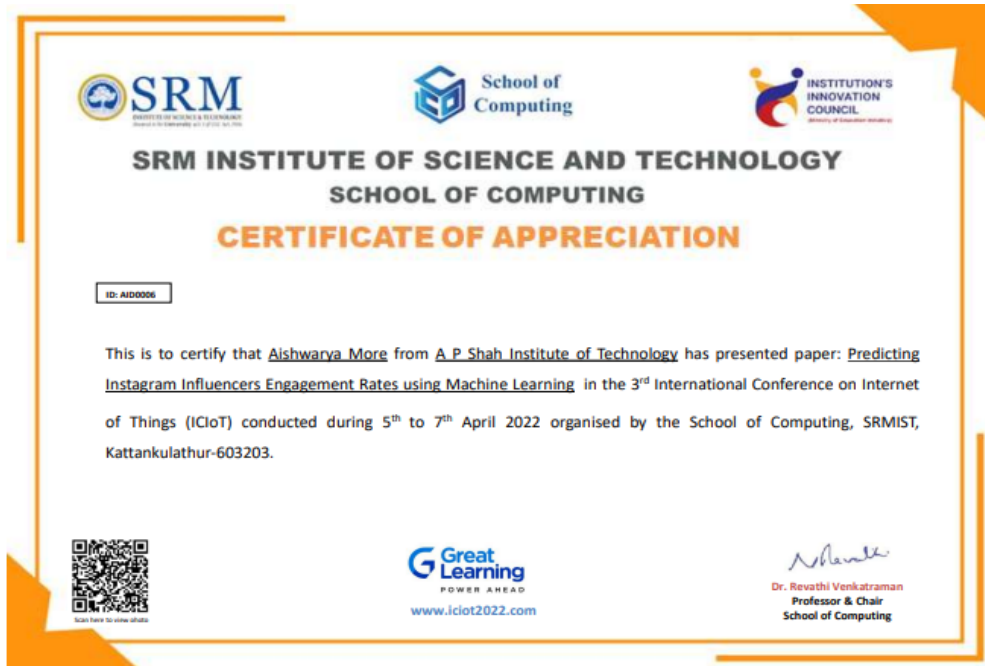
<https://youtu.be/iJGvYBH9mcY>

<https://youtu.be/WpSnulcRo8E>

<https://docs.python.org/3/library>

Chapter 8

Certificate





**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
SCHOOL OF COMPUTING**

CERTIFICATE OF APPRECIATION

ID: AID2009

This is to certify that Preksha Rajesh Vora from A.P Shah Institute of Technology has presented paper: Predicting Instagram Influencers Engagement Rates using Machine Learning in the 3rd International Conference on Internet of Things (ICIoT) conducted during 5th to 7th April 2022 organised by the School of Computing, SRMIST, Kattankulathur-603203.



Scan here to view photo



POWER AHEAD
www.iciot2022.com

Dr. Revathi Venkatraman
Professor & Chair
School of Computing



**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
SCHOOL OF COMPUTING**

CERTIFICATE OF APPRECIATION

ID: AID2008

This is to certify that Shah Moksha Paresh from A.P Shah Institute of Technology has presented paper: Predicting Instagram Influencers Engagement Rates using Machine Learning in the 3rd International Conference on Internet of Things (ICIoT) conducted during 5th to 7th April 2022 organised by the School of Computing, SRMIST, Kattankulathur-603203.



Scan here to view photo



POWER AHEAD
www.iciot2022.com

Dr. Revathi Venkatraman
Professor & Chair
School of Computing

Chapter 9

Log Book