

UE18CS390A - Capstone Project Phase - 1

SEMESTER - VI

END SEMESTER ASSESSMENT

Project Title : Forecasting future trending topics using text data.
Project ID : PW22DS03
Project Guide : Dinesh Singh
Project Team : Wise turtles
Members : Kavishankar K S PES1201802001
Mohammed Zeeshan PES1201801814
Rohan M PES1201801798

Problem Statement

Problem Statement:

Discovering trending topics from social media platforms and
Forecasting the future trends based on the recent and past trend of the
topics collected.

Abstract

Youtube , Google, and Twitter are often figures among the top 3 most popular application on website .

It has been reported Youtube users upload 72 hr videos / min.

So ,Forecasting content popularity is of great support and drive for design and management of various services.

our project is towards building ML model to forecast popular online content, using Various technique .

For this project we use past trending topic to train ML models for prediction future trendings.

Forecasting the future trending topics by using the recent and past trend data from social media sites like twitter, youtube and with the help of google trends.

Scope

Content Creators : Forecasting helps content creators to make a trending contents on their platforms like Youtube .

Social Analysis: Analysis of interests by region and its lifespan. Any kind of social analysis can help us gain an insight into the current psychology of the people. This can help us determine how the public will react to some event and make necessary preparations in advance.

Finance: Information related to companies and business magnates tend to surface on social networks like YouTube , Google and Twitter even before the official press release. These kinds of information can be useful in determining the stock trends and ultimately even to predict stock prices.

Marketing: People often share their opinions on anything and everything. This can be used to determine the interests of an individual or group. Based on these interests products can be targeted to specific audiences in order to maximize prot.

News Tracking: Information surfaces on YouTube , Google and Twitter very quickly. This allows news and the updates regarding it to be available instantaneously. Since These are a global platform, even international news can be updated in real-time.

Thus extracting trends from YouTube , Google and Twitter can prove to be extremely useful.

Literature survey

Paper 1: *Using Early View Patterns to Predict the Popularity of YouTube Videos*

- In this study **Multivariate Linear (ML) Model** , **MRBF Model** and **S-H Models** are used to predict future trends in youtube.
- By assigning different weights to different samples within the monitoring period, our model is able to better distinguish between videos with different patterns, which leads to significant reductions in average prediction errors.
- We also found that by exploring the similarity between the video and known the training set through RBF functions, we are able to reduce prediction errors even further, by adapting the prediction to better handle some specific kinds of videos.

Literature survey

Paper 2: Twitter Trends

- **MapReduce model** enhanced the speed of computing the vast amount of tweets and created a framework on which further analysis can be done.
- In our case, the (**Latent Dirichlet Allocation**) **LDA algorithm** is used to process a set of documents(collections in database) of tweets prepared after the extraction and pooling process.
- LDA provides us with a list of 'n' topics that constitute the documents. These topics are represented by a group of weighted words which can also be represented as a word cloud. Therefore, each word cloud can be considered as a trending topic.
- To improve the efficiency of the process, tweets containing the same entity are pooled together in the same document. The top 100 entities are considered for pooling.

Literature survey

Paper 3 : *Topic discovery and future trend forecasting for texts*

- They used predict the next trending research topics based on corpus of research paper collected.
- sentence-level pattern mining to discover meaningful topics.
- Preprocessing using association analysis and LSA.
- Used “WEKA” plugin which uses regression model for prediction.

Literature survey

Paper 4 : *Forecasting Time Series Data Using Hybrid GREY Neural Network And ARIMA Model.*

- In this study, a combination of GRANN ARIMA is proposed as a new approach for hybridizing linear and nonlinear models.
- The proposed model has few integrated features such as engaged with multivariate time series data.
- Forecasting accuracy of the proposed hybrid model, GRANN ARIMA is better compared to the individual model such as GRANN, ARIMA, MR and the second order error, LVM.

Suggestions from Review - 3

No major suggestion were given in review -3

Design Constraints

Design Constraints:

- As our project is based on real time data. For every result preprocessing and extraction of data is required.
- Since we are using ML models to predict the future trends, we may need huge time and high computational power if we use large datasets to train the model.
- A lot of time is consumed if we want to make changes to the model depending on the circumstances, since training the model requires a large amount of time.
- We are dealing with text data which is indeed a big constraint since we can't work on it, so we have to convert into some form before any process.
- Extent of forecasting.

Assumptions

Assumptions:

- The trending topics in future are based on the current and past trend.
- The frequency of occurrence of topics helps in predicting its trend.
- Accuracy of chosen model is high.
- Topics are low resolution and high sensitive.
- Rising and falling of the topic do not have a long term spanning.

Dependency

Dependency:

- Confounding variables in data.
Effects the both dependent and independent variable.
- Correlation between topics.
Extent of relation between each topic.
- Data Quality .
Large dataset and useful features in the dataset.
- Temporal dependency.
Casual relationship b\w the topics.

Design Details

- **For data extraction the Project depends on platform**
 - Twitter
 - Google trends
 - YouTube
- **Building User Interface.**
 - Flask API
- **Module and Libraries like**
 - Tensorflow and Keras
- **Platform for running and Testing.**
 - Jupyter notebook
 - Google Collab
- **Data Storage.**
 - Cloud storage.
- **Microservices for connecting process.**

Design Details

❖ **Novelty.**

The forecasting is used everywhere to progress, produce and provide things on demand.

Being able to tackle upcoming problems if they are initial known even if fairly.

❖ **Innovativeness.**

The project more innovative where we are going to forecast trends of various major fields and we are trying to predicting ramification may arise from upcoming trend.

❖ **Interoperability.**

System can able to exchange data with cloud storage and user for results and within the system.

Design Details

❖ **Maintainability**

Since the system is to be designed in fully automated way there where less maintainability.

❖ **Portability**

Highly portable and can be run in browser.

❖ **Legacy to modernization.**

It highly valuable wealth in the field of decision making, forecasting demands and precautions can be taken based on trend.

❖ **Performance**

Should be very high to interpret the results in real time.

Proposed Methodology / Approach

1. Data Extraction:

Web scraping

Using API's- provided by different social media platforms mainly from twitter, google trends and YouTube.

2. Preprocessing

Normalisation- remove noise from the data.

Removal of:

stop words: Commonly occurring words.

stemming: reducing word into root form.

3. Text Classification:

NLP: classifying the topic to domain it belongs

Pattern/rule based classifier.

SVM classifiers.

Word2vec.

Proposed Methodology / Approach

4. Future Selection for text classification:

Optimizing a model by selecting a subset of the features to use.

Reducing the overfitting of the dataset.

Wrapper method-Wrapping methods compute models with a certain subset of features and evaluate the importance of each feature.

5. Forecasting and modelling :

➤ Poison process:

variation of the particular topic over the time period.

➤ Markov chain:

trend analysis over period of time.

➤ Apriori algorithm- Conversion of transaction into frequent item set.

➤ Hybrid Model - for forecasting the trends.

➤ Clustering for violation detection.

Proposed Methodology / Approach

6. Testing

RMSE:

The model accuracy is evaluated using the RMSE.

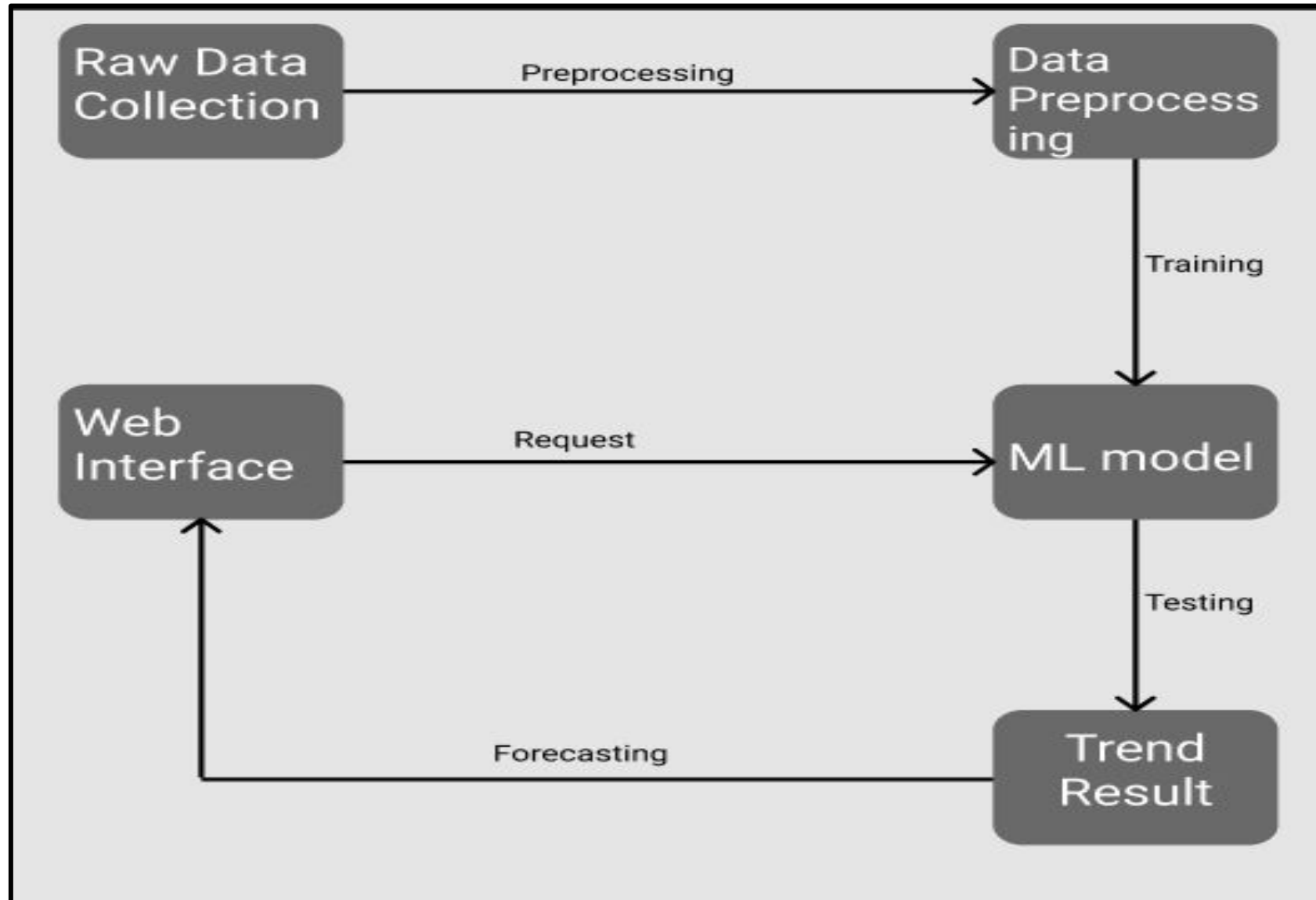
$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - F_t)^2}$$

7. Interpreting the result.

Visual graphs.

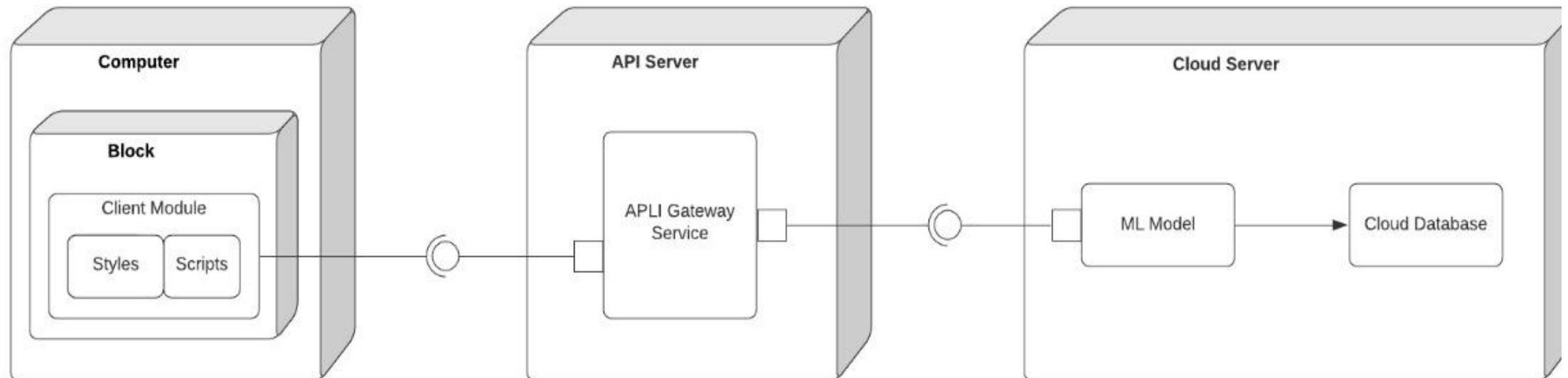
Future Trend Score.

Architecture

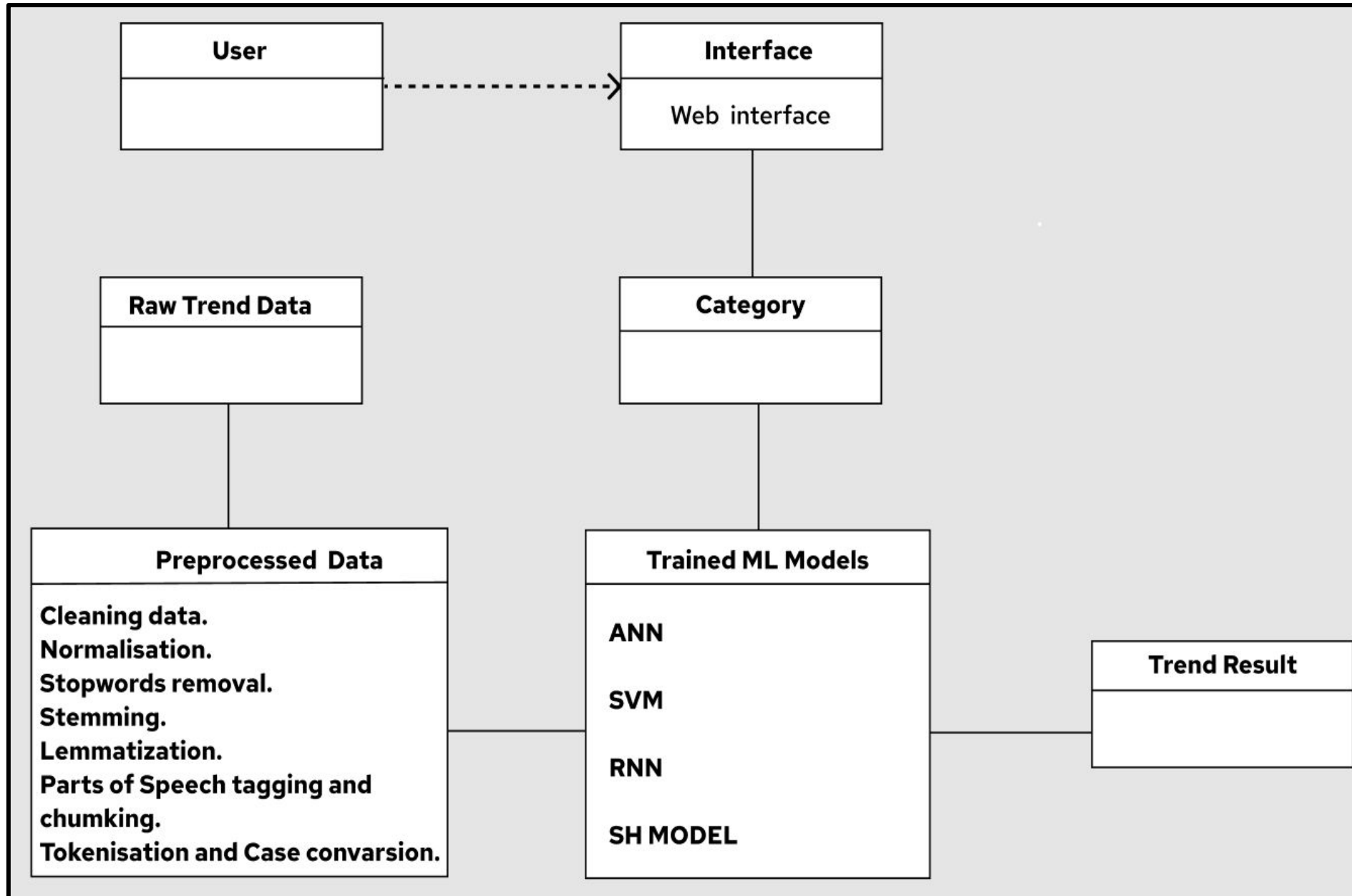


Architecture

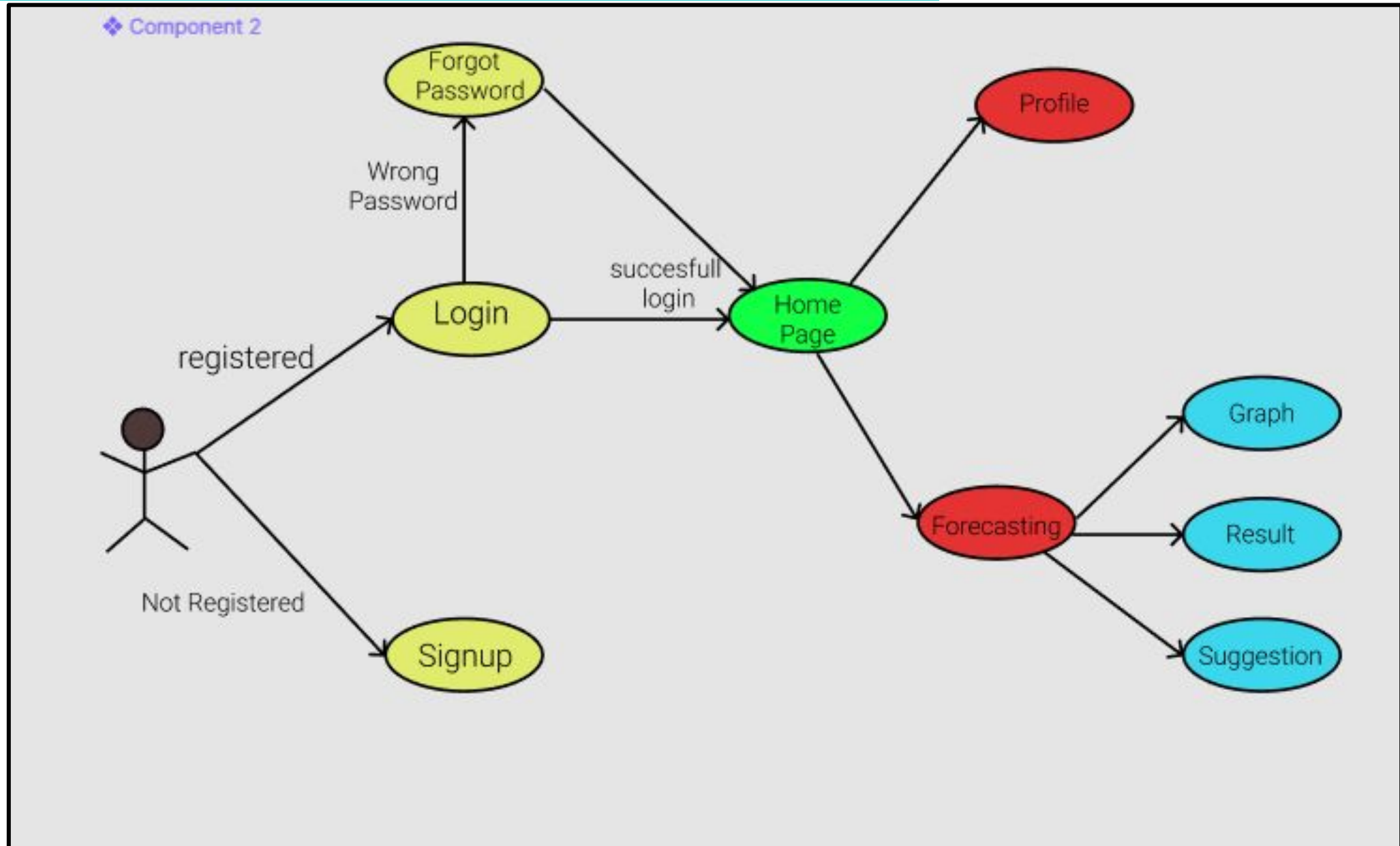
DEPLOYMENT DIAGRAM



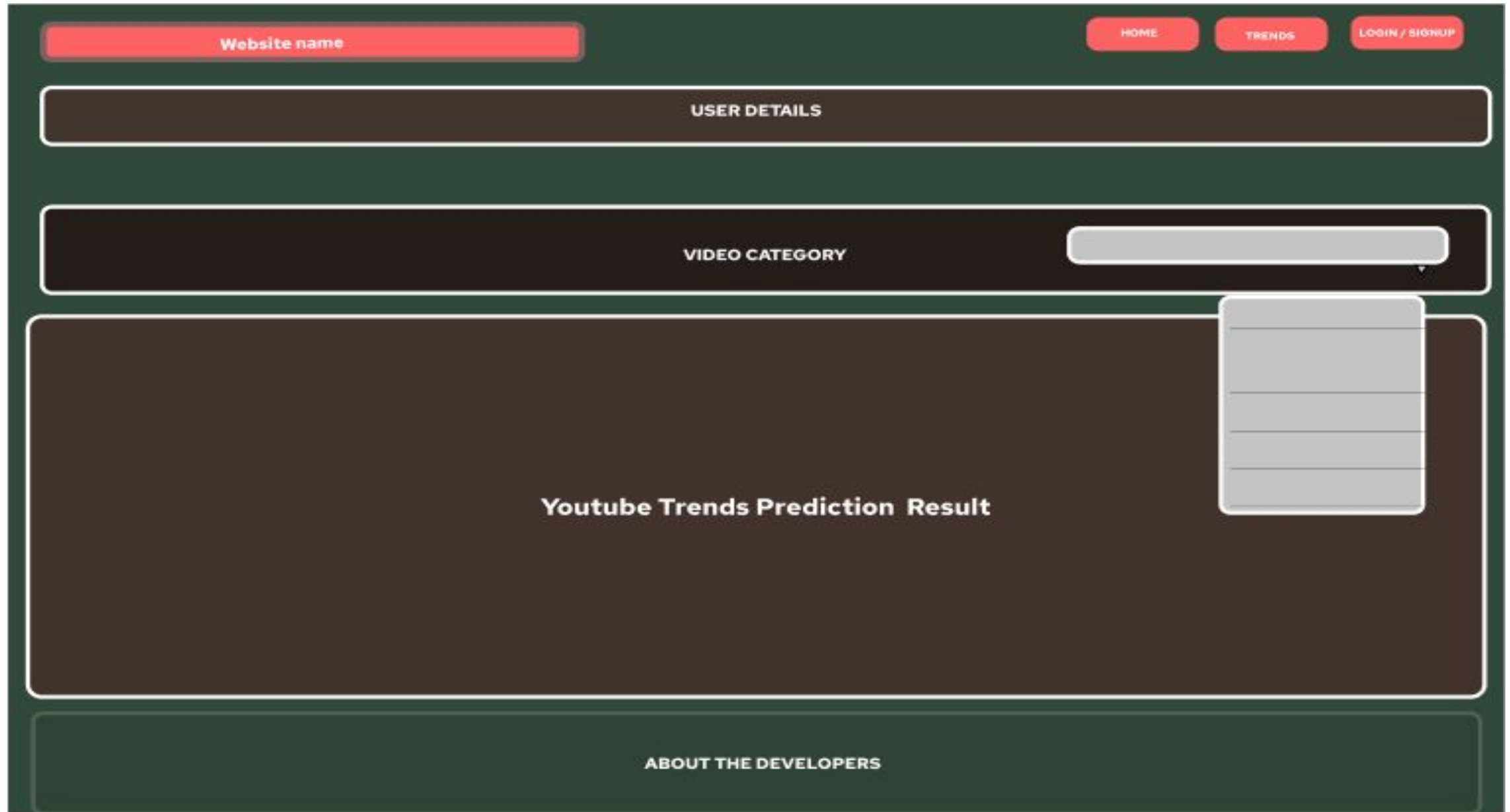
Master Class Diagram



Use Case Diagram



User Interface



Technologies Used

❖ Data Extraction:

- Using the web scraping.
Python BeautifulSoup.
- Using the api's provided by specific social media websites.
 - Twitter Developer API
 - Youtube API
 - Google Trends API

❖ Data preprocessing:

- Using NLP techniques.
 - Normalisation.
 - Stopping and Stemming.
 - POS Tagger.
 - Lemmatisation

Technologies Used

❖ Forecasting:

- Methodes
 - Statistical mode
 - Artificial Intelligence
- Data analytics techniques.
 - ARIMA
 - MR
- Hybrid model
 - GRAN-ARIMA.

❖ Visualization tools:

- Generating real time graphs.
Python Libraries like matplotlib and Pyplot.
- WordCloud
Generating cloud of words.

Technologies Used

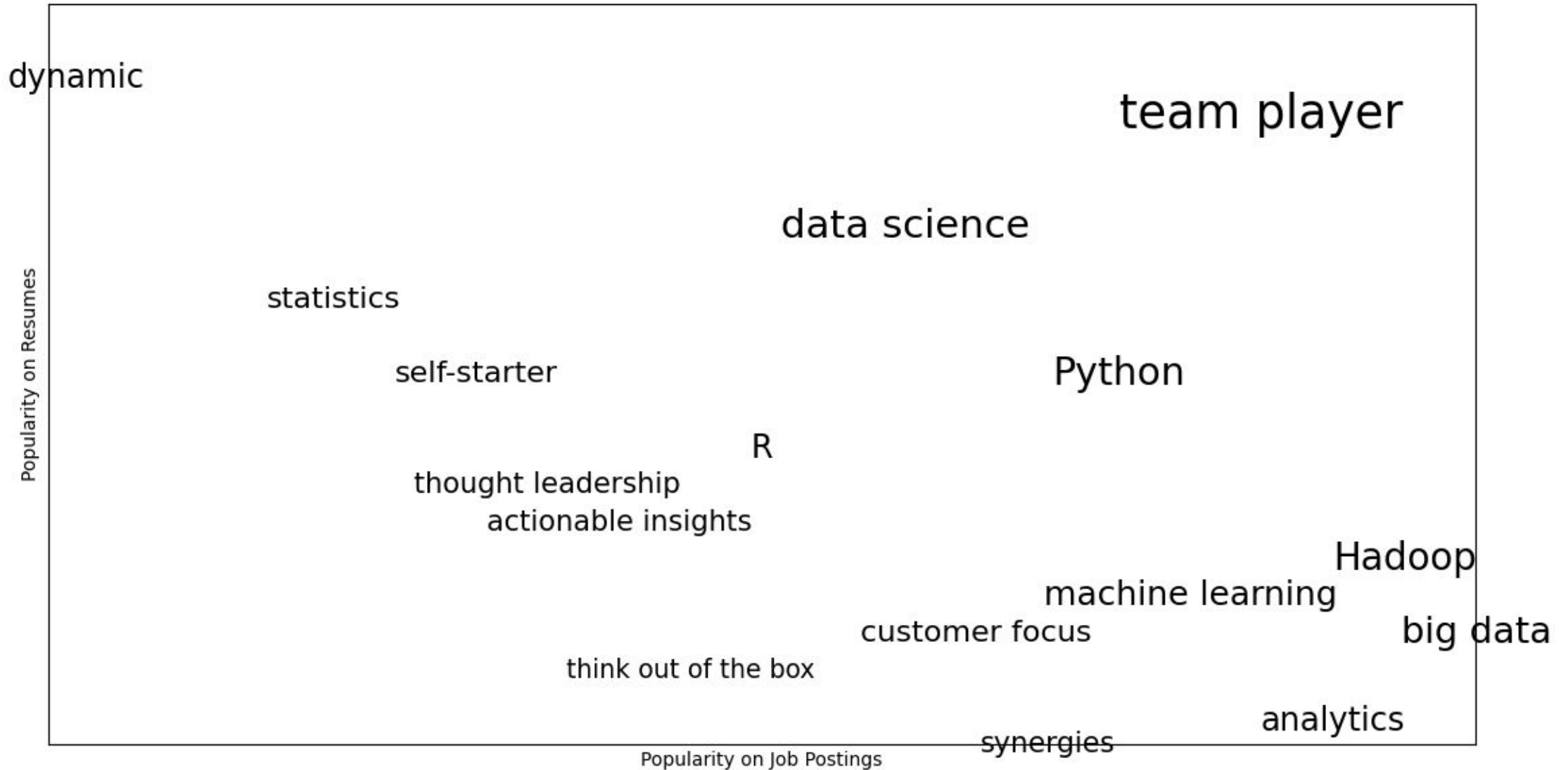
- ❖ **Flask:**
 - Web application
 - Machine learning model deployment.
- ❖ **Data Storage**
 - Storing data in cloud.
- ❖ **Testing**
 - Pycharm.
 - Jupyter.
 - Google Colab.

Project Progress

1. **Data collection using web scraping and API's of social networking platforms.**
2. **Preprocessing of data.**
3. **Currently working on choosing and building Machine Learning Model,based on the literature survey .**
4. **Analysis of Gathered data.**

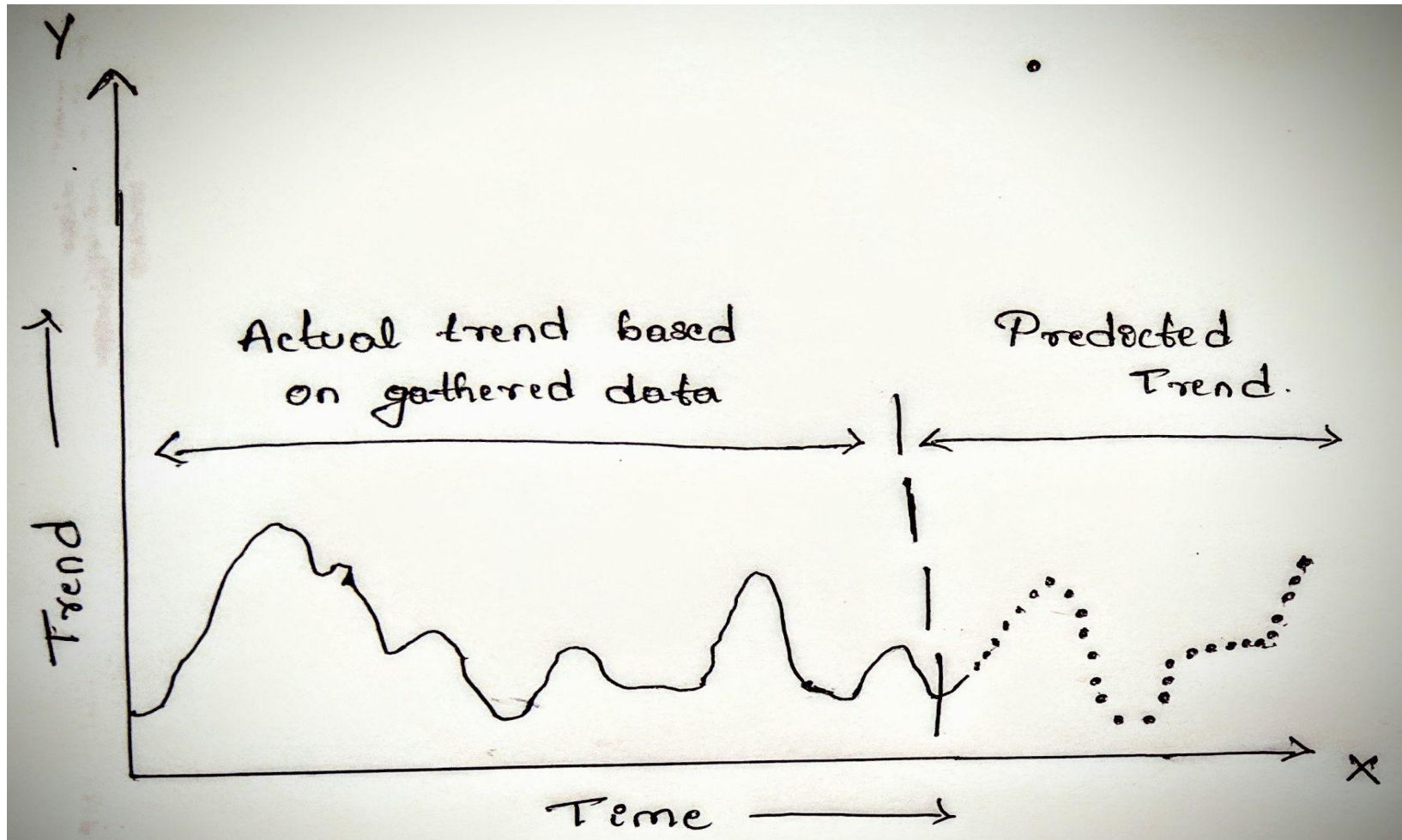
About 30% of the project has been completed till now.

Word Cloud :



Final interpretation of result

Available data is plotted in solid line and predicted data is plotted in dotted line.

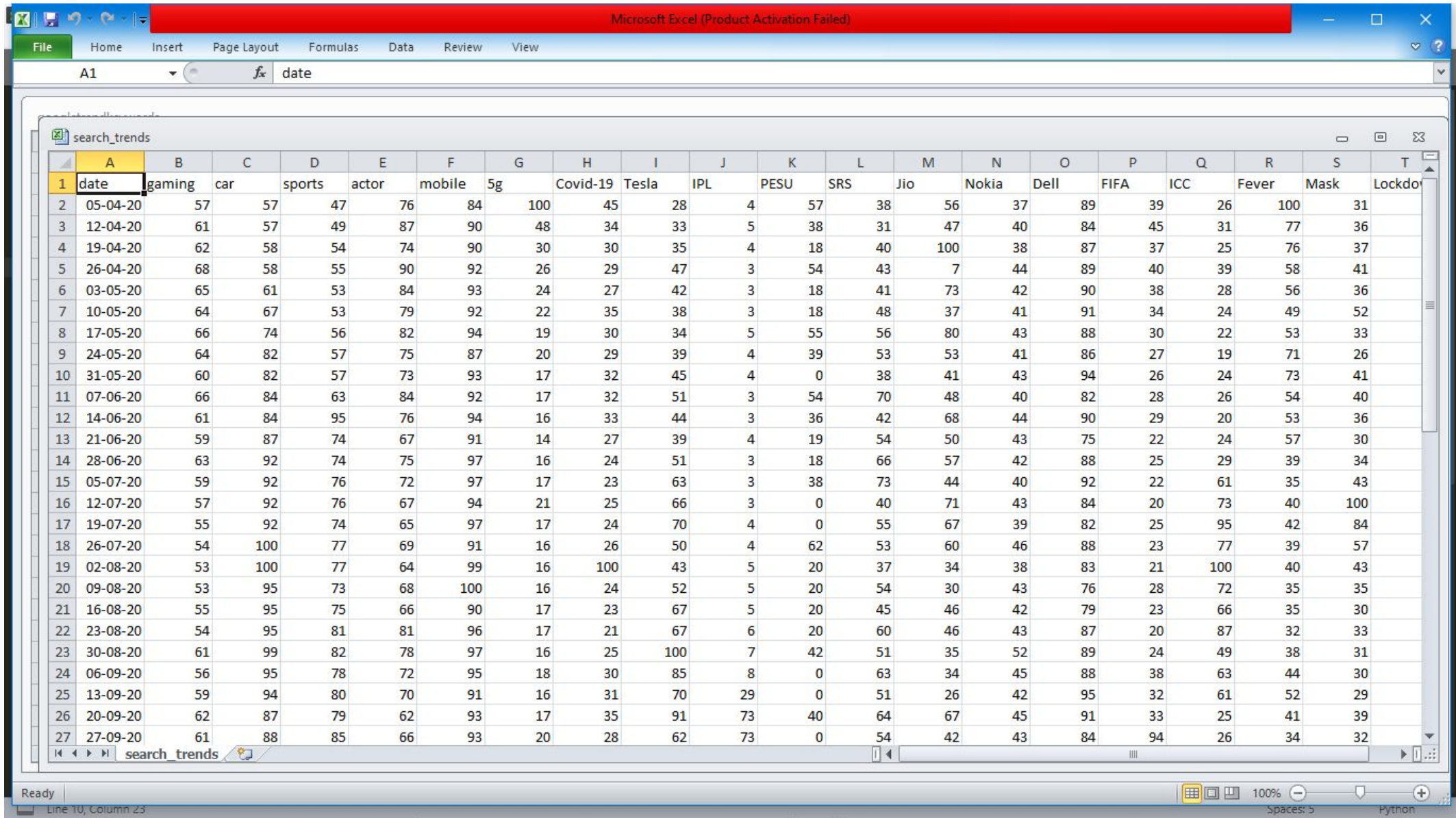


Demo

Pseudocode :

- 1.Extract data from social networks using web scraping and api's.
- 2.Preprocess Data using NLP and Normalisation ,Removal of POS, stop words, stemming etc.
- 3.Using extracted Data build a model for training
- 4.Train the model.
- 5.Test the trained Model and update the model as required.

Search Data Extracted From Google Trends using API :



Microsoft Excel (Product Activation Failed)

File Home Insert Page Layout Formulas Data Review View

A1 fx date

search_trends

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	date	gaming	car	sports	actor	mobile	5g	Covid-19	Tesla	IPL	PESU	SRS	Jio	Nokia	Dell	FIFA	ICC	Fever	Mask	Lockdo
2	05-04-20	57	57	47	76	84	100	45	28	4	57	38	56	37	89	39	26	100	31	
3	12-04-20	61	57	49	87	90	48	34	33	5	38	31	47	40	84	45	31	77	36	
4	19-04-20	62	58	54	74	90	30	30	35	4	18	40	100	38	87	37	25	76	37	
5	26-04-20	68	58	55	90	92	26	29	47	3	54	43	7	44	89	40	39	58	41	
6	03-05-20	65	61	53	84	93	24	27	42	3	18	41	73	42	90	38	28	56	36	
7	10-05-20	64	67	53	79	92	22	35	38	3	18	48	37	41	91	34	24	49	52	
8	17-05-20	66	74	56	82	94	19	30	34	5	55	56	80	43	88	30	22	53	33	
9	24-05-20	64	82	57	75	87	20	29	39	4	39	53	53	41	86	27	19	71	26	
10	31-05-20	60	82	57	73	93	17	32	45	4	0	38	41	43	94	26	24	73	41	
11	07-06-20	66	84	63	84	92	17	32	51	3	54	70	48	40	82	28	26	54	40	
12	14-06-20	61	84	95	76	94	16	33	44	3	36	42	68	44	90	29	20	53	36	
13	21-06-20	59	87	74	67	91	14	27	39	4	19	54	50	43	75	22	24	57	30	
14	28-06-20	63	92	74	75	97	16	24	51	3	18	66	57	42	88	25	29	39	34	
15	05-07-20	59	92	76	72	97	17	23	63	3	38	73	44	40	92	22	61	35	43	
16	12-07-20	57	92	76	67	94	21	25	66	3	0	40	71	43	84	20	73	40	100	
17	19-07-20	55	92	74	65	97	17	24	70	4	0	55	67	39	82	25	95	42	84	
18	26-07-20	54	100	77	69	91	16	26	50	4	62	53	60	46	88	23	77	39	57	
19	02-08-20	53	100	77	64	99	16	100	43	5	20	37	34	38	83	21	100	40	43	
20	09-08-20	53	95	73	68	100	16	24	52	5	20	54	30	43	76	28	72	35	35	
21	16-08-20	55	95	75	66	90	17	23	67	5	20	45	46	42	79	23	66	35	30	
22	23-08-20	54	95	81	81	96	17	21	67	6	20	60	46	43	87	20	87	32	33	
23	30-08-20	61	99	82	78	97	16	25	100	7	42	51	35	52	89	24	49	38	31	
24	06-09-20	56	95	78	72	95	18	30	85	8	0	63	34	45	88	38	63	44	30	
25	13-09-20	59	94	80	70	91	16	31	70	29	0	51	26	42	95	32	61	52	29	
26	20-09-20	62	87	79	62	93	17	35	91	73	40	64	67	45	91	33	25	41	39	
27	27-09-20	61	88	85	66	93	20	28	62	73	0	54	42	43	84	94	26	34	32	

search_trends

Ready Line 10, Column 23 100% Spaces: 5 Python

Keyword Used

A20		fx		Lockdown	
googletrendkeywords					
	A	B	C	D	
1	Keywords				
2	gaming				
3	car				
4	sports				
5	actor				
6	mobile				
7	5g				
8	Covid-19				
9	Tesla				
10	IPL				
11	PESU				
12	SRS				
13	Jio				
14	Nokia				
15	Dell				
16	FIFA				
17	ICC				
18	Fever				
19	Mask				
20	Lockdown				
21					

Search Data Extracted From Youtube using API :

21.23.04_IN_videos.csv - Excel

Rohan M

Share

FileHomeInsertPage LayoutFormulasDataReviewViewHelpTell me what you want to do

CutCopyFormat Painter

Clipboard

Calibri11

Font

Wrap Text

Alignment

Merge & Center

Alignment

General

Number

Conditional Formatting

Styles

Cell Styles

Styles

Insert

Cells

Delete

Cells

Format

Cells

Σ AutoSum

Editing

Fill

Editing

Clear

Editing

Sort & Filter

Editing

Find & Select

Editing

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	video_id	title	publishedAt	channelId	channelTitle	categoryId	trending	tags	view_count	likes	dislikes	comment	thumbnail	comments	ratings	dislike	description		
2	zPI7y5Ybzuo	Radhe: Your Most Wanted Bhai Official Trailer	2021-04-22T05:30	UCZ3MqC_Salman Khan Films	24	21.23.04	Salman Kh	6592573	480619	39260	72825	https://i.yt	FALSE	TRUE	FALSE		This Eid, there will be action, drama, m		
3	hAxqygRdM4g	Earth Day 2021 Doodle	2021-04-21T11:00	UCdq61mGoogleDoodles	1	21.23.04	[none]	0	0	0	0	https://i.yt	TRUE	FALSE	TRUE		This yearâ€™s annual Earth Day Doodl		
4	WeWXC_ivl5s	TVF's Aspirants Web Series Episode 3	2021-04-21T07:30	UCNjCSUS The Viral Fever	24	21.23.04	the viral fe	5574753	481107	6106	38376	https://i.yt	FALSE	FALSE	FALSE		Like the Episode? Rate us on IMDB - ht		
5	Jmot0qTEwEY	Kanimangalam kovilakam SEASON 2	2021-04-22T14:34	UCvioWRCLAPBOARD FILMS	24	21.23.04	#kanimang	613405	82958	814	5050	https://i.yt	FALSE	FALSE	FALSE		Mmuchstac Skin Lightening Oil and Ocea		
6	FhmK9RBkF64	Muh Dikhayi the mridul Pragati Niti	2021-04-21T13:15	UC0zXnLIF The MriDul	23	21.23.04	the mridul	3725393	160257	6853	7156	https://i.yt	FALSE	FALSE	FALSE		Hello Doston ðŸ‘¸ðŸ’ŠðŸ”ŠðŸŽ‰ðŸŒ©Fç		
7	I6Emecdkgc	EMIWAY - SHY (PROD. MEME MACHINE)	2021-04-22T11:27	UCHmk8iN Emiway Bantai	10	21.23.04	emiway e	2405733	409750	8592	27156	https://i.yt	FALSE	FALSE	FALSE		Artist - Emiway Bantai Music - Meme M		
8	27qgWtUkyw	Letting YOUTUBER'S decide a DARE for	2021-04-21T07:16	UCIfos9f7i Fukra Insaan	22	21.23.04	fukra insaa	2439346	375054	5162	14787	https://i.yt	FALSE	FALSE	FALSE		Open your Free Upstox Gold Account a		
9	8ssSaWfB9Rk	Surya Web Series Episode - 6 Shan	2021-04-20T13:30	UCjBFvKC Shanmukh Jaswanth	24	21.23.04	shanmukh	6066121	354715	5919	27510	https://i.yt	FALSE	FALSE	FALSE		Download the OctaFX Copytrading App		
10	5NdUPDZFhyt	Tum Pe Hum Toh Bole Chudiyen Nawa	2021-04-22T05:30	UCFfbwnv Zee Music Company	10	21.23.04	Tum Pe Hu	2987960	162791	4334	5576	https://i.yt	FALSE	FALSE	FALSE		ðŸŒ™ðŸŒ™ SUBSCRIBE to Zee Music Con		
11	KbOxmDmoTs	Bakery Alaparaigal #Nakkalites	2021-04-21T12:55	UCipmhfdi Nakkalites	23	21.23.04	Bakery Ala	1022193	70964	3183	2841	https://i.yt	FALSE	FALSE	FALSE		Open your Free Upstox Gold Account a		
12	Y6Ta6LC9bm8	Eruma Saani MOM vs SON	2021-04-21T14:54	UCPyFYIqb Eruma Saani	23	21.23.04	Eruma Saa	765228	75067	826	2108	https://i.yt	FALSE	FALSE	FALSE		Open your Free Upstox Gold Account a		
13	plEcdu4kKMc	Unique Mud Fishing Using Coca Cola & E	2021-04-21T05:32	UCxGWGP Travel Master	24	21.23.04	survival ski	3759948	357374	8828	22286	https://i.yt	FALSE	FALSE	FALSE		unique mud fishing Using coca cola & e		
14	ubTVVmJIZIs	Pregnancy Girls Zone Ft Ival Nandhini,	2021-04-21T13:04	UCKoQ57V Unakkennapa	24	21.23.04	unakkenna	546705	42928	472	1254	https://i.yt	FALSE	FALSE	FALSE		#pregnancy #girlszone #unakkennapa		
15	iWLyyQE-AEQ	NILA'S ARRIVAL PEARLE MAANEY SR	2021-04-19T05:41	UCeAlITNJ Pearle Maaney	24	21.23.04	pearle ma	4182111	259666	3536	18223	https://i.yt	FALSE	FALSE	FALSE		Poster by : Rahul OzLike me on Facebo		
16	kfhucicNALY	Enga Veetu Kalyanam Finale Its VG	2021-04-21T11:30	UCzKchRia Its VG	26	21.23.04	Enga Veetu	765609	34752	677	1053	https://i.yt	FALSE	FALSE	FALSE		My skin care routine includes.Dermaviv		
17	OnjB2K5rbiE	Star Magic Flowers EP#340	2021-04-20T09:43	UCmApdlr Flowers Comedy	24	21.23.04	flowers tv	2710145	78048	3202	5603	https://i.yt	FALSE	FALSE	FALSE		90+ à'tà'âmà'zà'zà'âmà'•àm#à'à'à'a'		
18	SsPrdSM_SOc	Desi Mom vs Modern Mom ShrutiArjun	2021-04-20T12:00	UCIsotCCc Shruti Arjun Anand	24	21.23.04	shruti shru	2587670	186642	9439	4294	https://i.yt	FALSE	FALSE	FALSE		Maa sabki ek si hoti hai... Lekin unki kh		
19	Rh5XqnzYPQC	Wow Life Presents Achumaâ€™s Fridge	2021-04-21T08:30	UC547TD5 Wow Life	24	21.23.04	wow life e	959121	24081	2787	1732	https://i.yt	FALSE	FALSE	FALSE		A fridge is almost the most important e		
20	H49CTeac58U	GROWING OP TOOLS FOR LOGGY BIRTH	2021-04-22T07:00	UC3MTnx5 Chapati Hindustani Ga	20	21.23.04	minecraft	1875506	193024	2658	24427	https://i.yt	FALSE	FALSE	FALSE		GROWING OP TOOLS FOR LOGGY BIRT		
21	0FZSPwXi3kU	Thondara Padaku Sundara Vadhana E	2021-04-22T12:32	UCcAOWk Don Pruthvi	24	21.23.04	Pruthvi Th	705044	53723	788	4713	https://i.yt	FALSE	FALSE	FALSE		To Order Jampani PicklesWhatsapp: ht		
22	bRvDKVko7zk	Kalyanam to Kaadhal Episode 1 Mini V	2021-04-22T11:15	UCPsy4hW JFW - Just for Womer	24	21.23.04	Jfw Jfw w	210578	11360	148	433	https://i.yt	FALSE	FALSE	FALSE		Kalyanam to kaadhal EPISODE 1 Intro		
23	eEDR1yJ9ymk	Maharashtra Lockdown : à¤†à¤œ à¤†à¤œ	2021-04-22T03:39	UCrcpw88 News18 Lokmat	25	21.23.04	Maharashi	315299	1368	222	136	https://i.yt	FALSE	FALSE	FALSE		#Lockdown #Coronalatestupdate #Ma		
24	Bk1-ERw1Wo	We All Locked Up In Lockdown ðŸ’Š...	2021-04-22T16:55	UC027uEe Mohit Chhikara	24	21.23.04	mohit mo	328344	52705	406	1364	https://i.yt	FALSE	FALSE	FALSE		We all locked up in our office so we car		
25	OdJ9fpBMTz4	Ram Leela Web Series Episode - 1	2021-04-21T12:32	UCVG75EN Hey Siri	24	21.23.04	siri hanum	659042	45691	757	2321	https://i.yt	FALSE	FALSE	FALSE		Open your Free Upstox Gold Account a		
26	JdBYVNuky1N	Apple Event â€œ April 20	2021-04-20T18:05	UCE_M8A Apple	28	21.23.04	Apple App	8631506	325878	8791	0	https://i.yt	TRUE	FALSE	FALSE		Watch the special Apple Event and lear		
27	_pGFmV85dKI	Finally, The Wait is OVERRR !!! Check D	2021-04-22T13:30	UCIGyWNI Physics Wallah - Alakt	27	21.23.04	class 11 p	322070	29104	507	4157	https://i.yt	FALSE	FALSE	FALSE		Welcome To Class 11th : https://youtu		
28	tKWwF2XpH	Goriyaan Goriyaan Romaana ft Jasmin	2021-04-21T09:31	UC783dzn DM - Desi Melodies	10	21.23.04	goriyaan g	5873257	86147	2177	8182	https://i.yt	FALSE	FALSE	FALSE		The boy next door Romaana a promisi		
29	DhiNMAduqx	Ritik Left Tsg ? - Where Is He Full Reaso	2021-04-22T06:17	UCiDiKVE7 TWO SIDE GAMERS	20	21.23.04	Two side G	1427448	303633	3272	23519	https://i.yt	FALSE	FALSE	FALSE		Be Safe & Stay Home #Twosidegamers		

21.23.04_IN_videos

100%

01:48 PM 23-04-2021

Data extracted from twitter trends :

```
C:\Users\Zeeshan\Desktop\capstone\r3 ppt script - Sublime Text (UNREGISTERED)

Command Prompt

Hashtag: Daniel Kaluuya      volume of tweets: 120587
Hashtag: Harrison Ford      volume of tweets: 10539
Hashtag: Trent Reznor       volume of tweets: 10283
Hashtag: Jessica            volume of tweets: 63528
Hashtag: Sound of Metal     volume of tweets: 48492
Hashtag: Don Sergio         volume of tweets: 48411
Hashtag: Chlo  Zhao         volume of tweets: 125131
Hashtag: Globo              volume of tweets: 116207
Hashtag:                   volume of tweets: None
Hashtag: bbb18              volume of tweets: 33031
Hashtag: Corinthians        volume of tweets: 88744
Hashtag: Pocah e Arthur     volume of tweets: 63646
Hashtag: Fernando Carlos    volume of tweets: None
Hashtag: Kershaw            volume of tweets: None
Hashtag: Steven Yeun        volume of tweets: 20134
Hashtag: Aranza             volume of tweets: 16211
Hashtag: Agora o Arthur     volume of tweets: 34387
Hashtag: Riz Ahmed          volume of tweets: 27262
Hashtag: Artur              volume of tweets: 14842
Hashtag: Elana              volume of tweets: 15623

C:\Users\Zeeshan\Desktop>
```

Summary of work done in Capstone Phase-1

- Collection of research paper in this subject- Jan/21
- Analysis of Collected research paper along with team-Feb/21
- Formalisation of extraction data using API and Finding API's Feb/21
- Data Extraction using API and Discussion on model for forecasting -Mar/21
- Discussion on interpreting results and microservices - Apr/21
- Discussion on what additional things can be added -Apr/21
- Preparing to make project to tackle business problems -Apr/21

Project Plan for Capstone Phase-2

- Finding the method to organize all the extracted data.
- Extensive Preprocessing of the collected Data.
- Extraction and Classification of the topics from data.
- Feature selection from the data.
- Building models for forecasting of text data.
- Training and Evaluating model.
- Building a user interface.

References

References:

1. Jose L. Hurtado*† , Ankur Agarwal† and Xingquan Zhu†,
"Topic discovery and future trend forecasting for texts"
2. Roselina Sallehuddin, Siti Mariyam Hj. Shamsuddin, Siti Zaiton Mohd. Hashim,,Ajith Abrahamy, "Forecasting Time Series Data Using Hybrid GREY Neural Network And ARIMA Model".

References

3. “YouTube View Prediction with Machine Learning”

Int. J. Business Information Systems, Vol. 13, No. 3, 2013 359

4. Conrad Tucker¹ and Harrison M. Kim¹ (1)

“PREDICTING EMERGING PRODUCT DESIGN TREND BY MINING PUBLICLY AVAILABLE CUSTOMER REVIEW DATA”.- University of Illinois at Urbana-Champaign, USA.

**Thank
You**