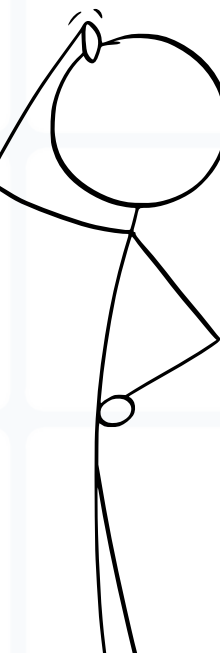
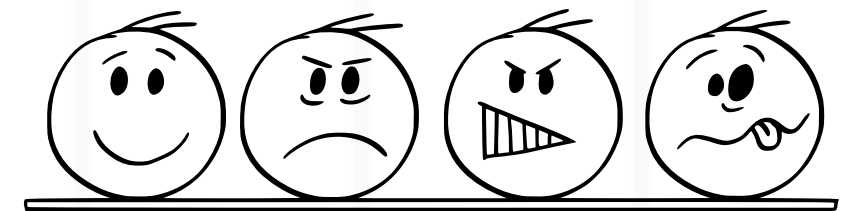


..... **MLPR PROJECT**

PLAKSHA HAPPINESS INDEX

aaj khush toh bohot hoge tum...shayad



Shaurya Sighadia
Suhani Jain

THE PROBLEM

SEM I

Bob is full of energy

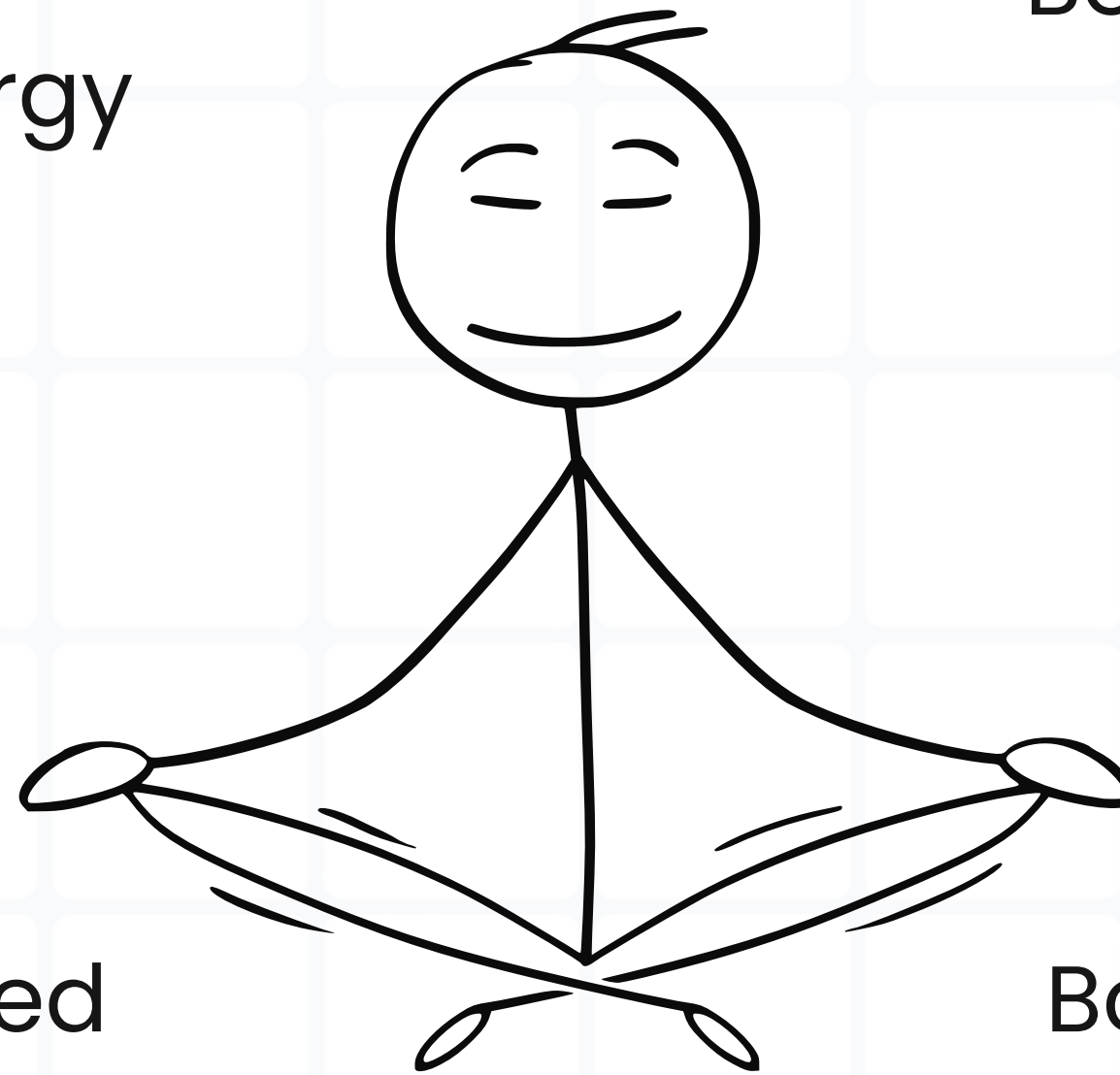
Bob is at the top of his class

Bob has a zeal for life

Bob has good focus

Bob is happy and motivated

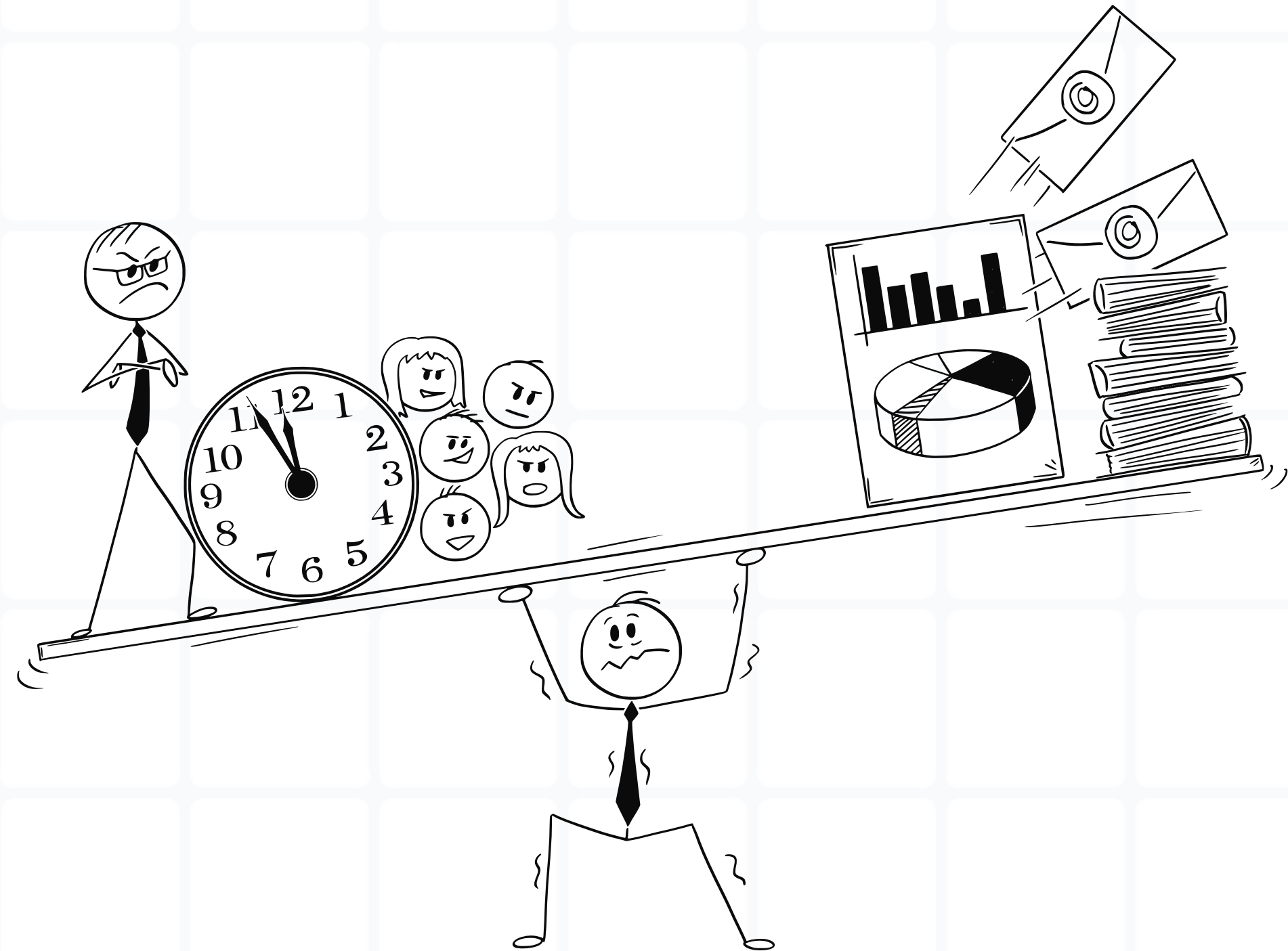
Bob is living



THE PROBLEM

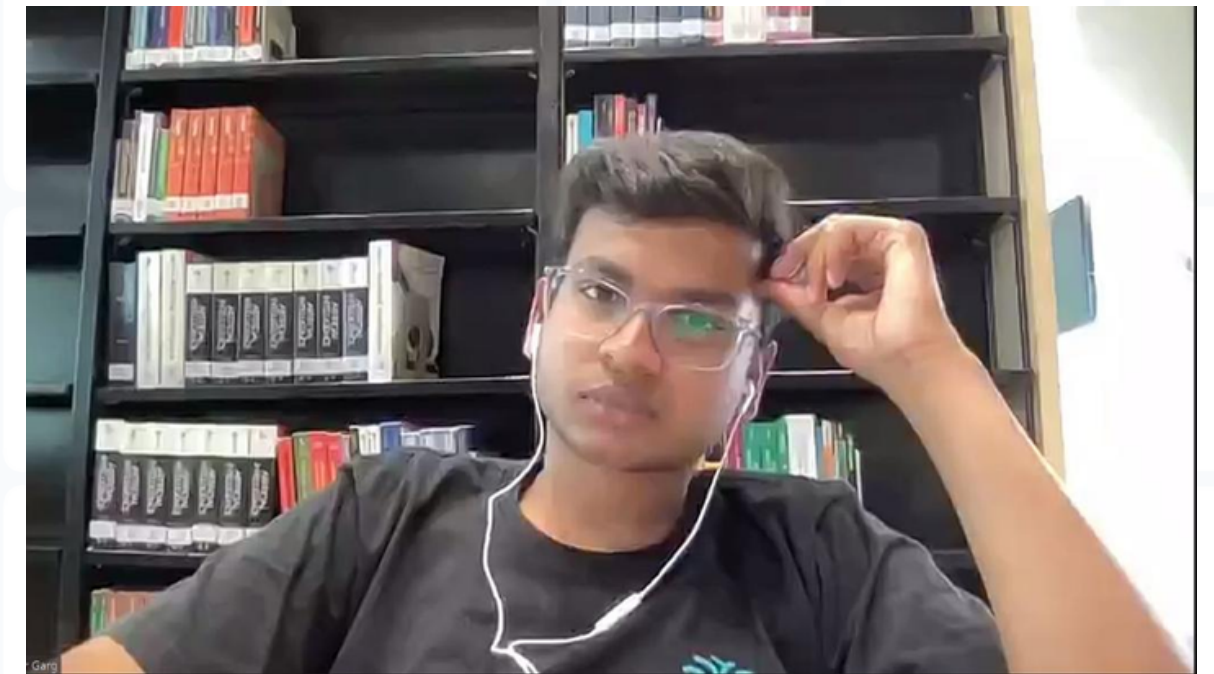
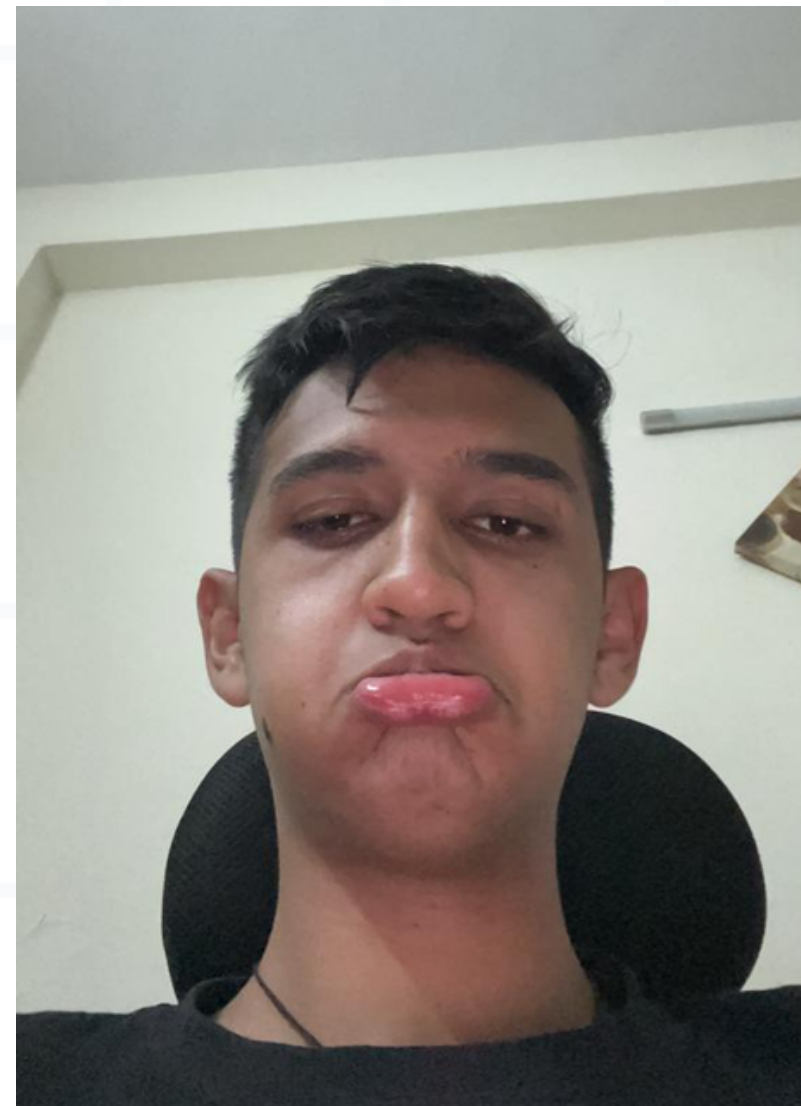
SEM V

- Increased stress
- Mental Exhaustion
- Decreased focus and motivation
- Pressure of failure
- Low Happiness quotient



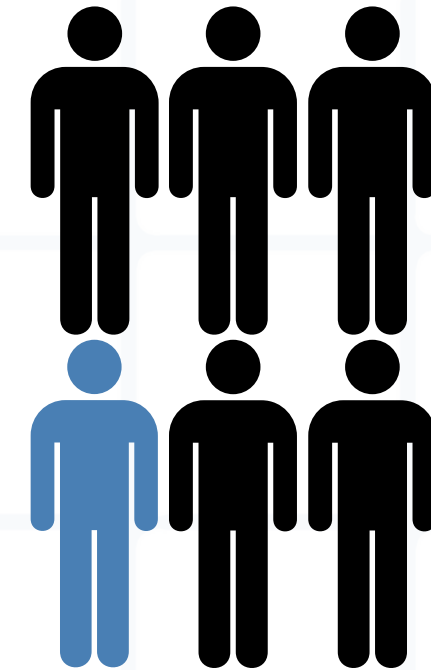
THE PROBLEM

True Students at Plaksha facing
the same problems causing
increased mental stress



THE PROBLEM

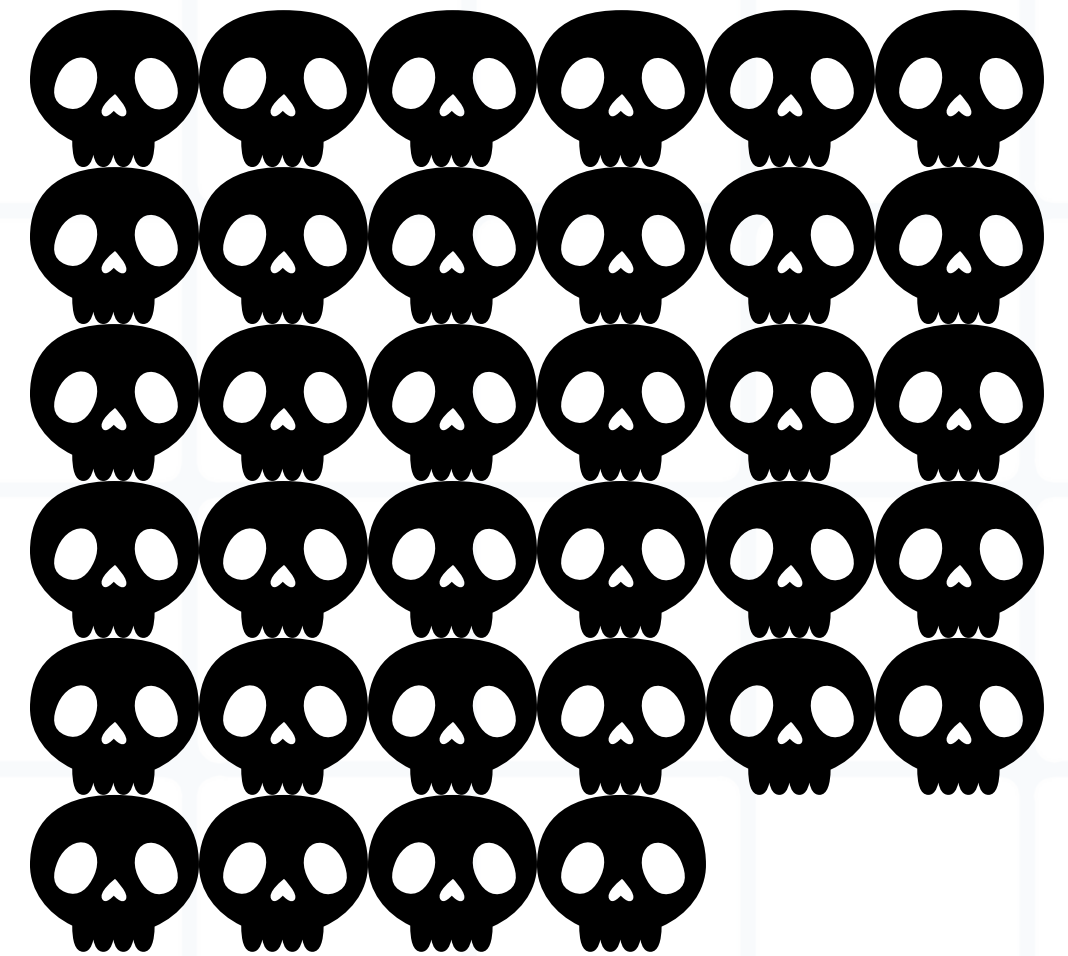
The frequency of depression, anxiety and stress among university students was found 75%, 88.4% and 84.4% respectively.



THE PROBLEM

More than 34 students in India attempt suicide every day due to academic stress.

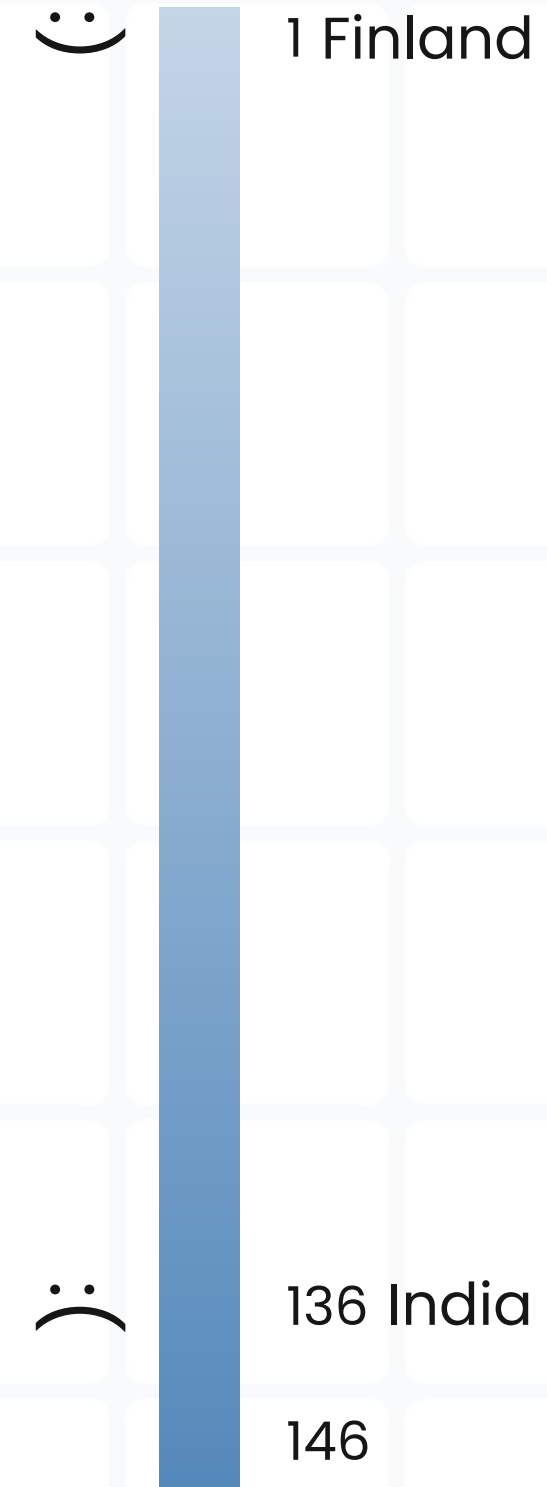
[Source](#)



THE PROBLEM

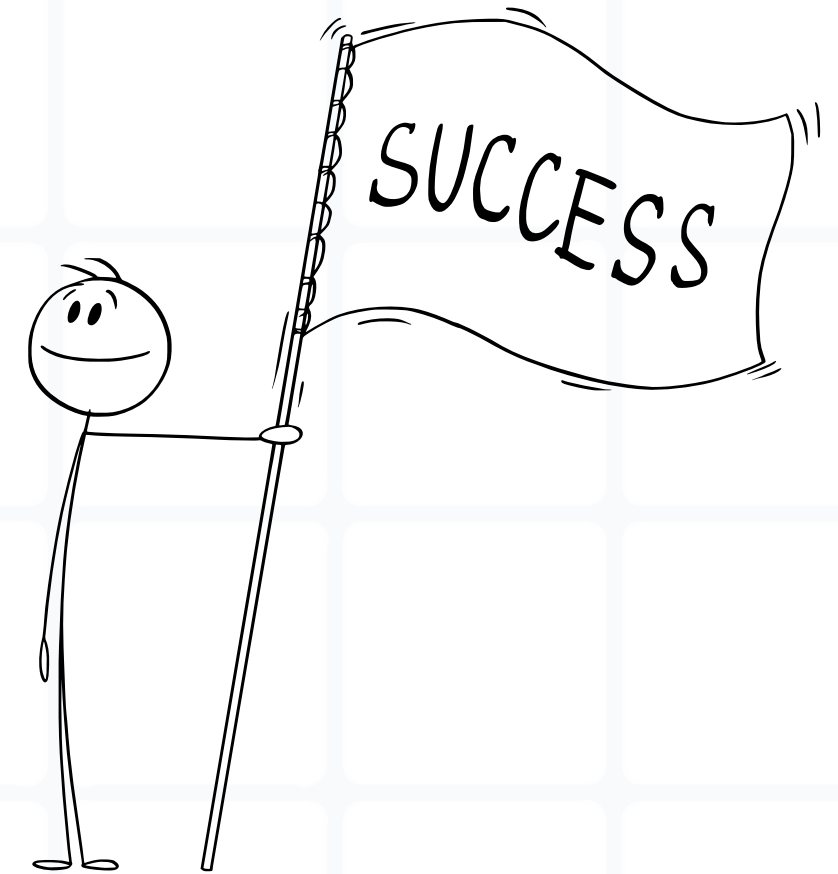
According to the 2022 World Happiness Report, India is in the 136th position among 146 countries

[Source](#)



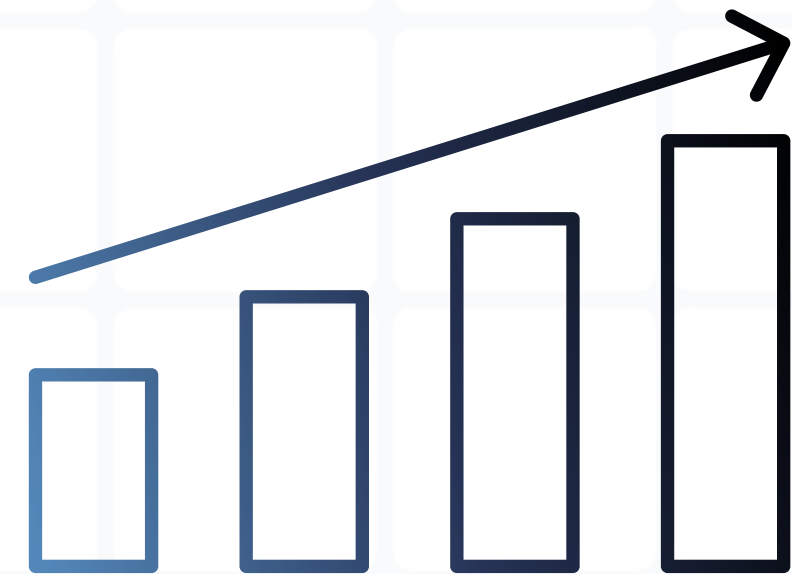
MOTIVATION

There is a direct correlation
between student happiness
and academic achievement



MOTIVATION

A happy student has better memory, retention and focus than an unhappy student

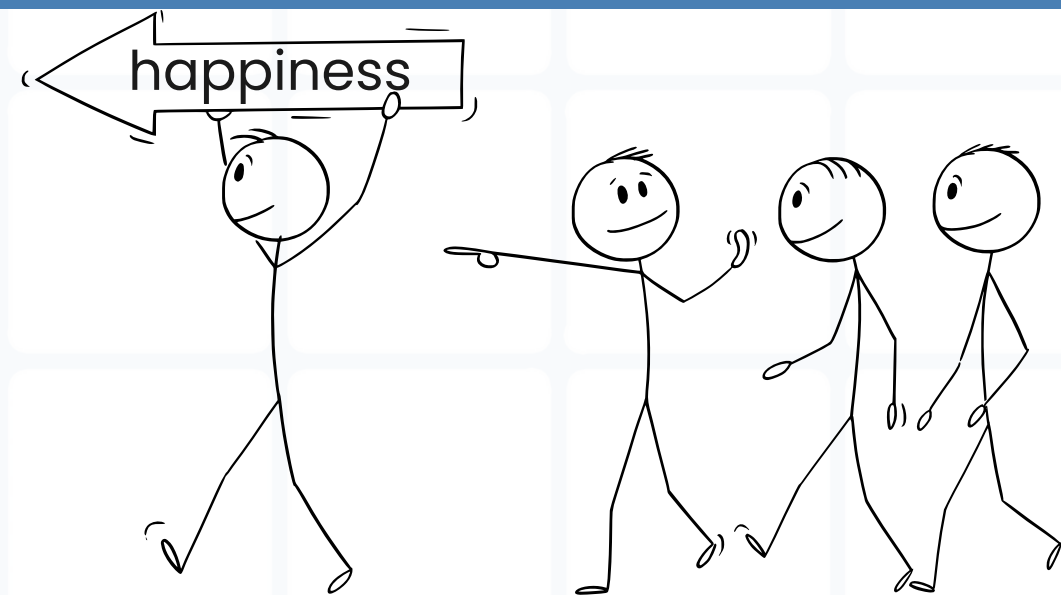


OUR GOAL

“

Create the Plaksha Happiness Index (PHI) and identify the factors that contribute most towards happiness at Plaksha, using insights to provide relevant recommendations to the administration to improve student life.

”

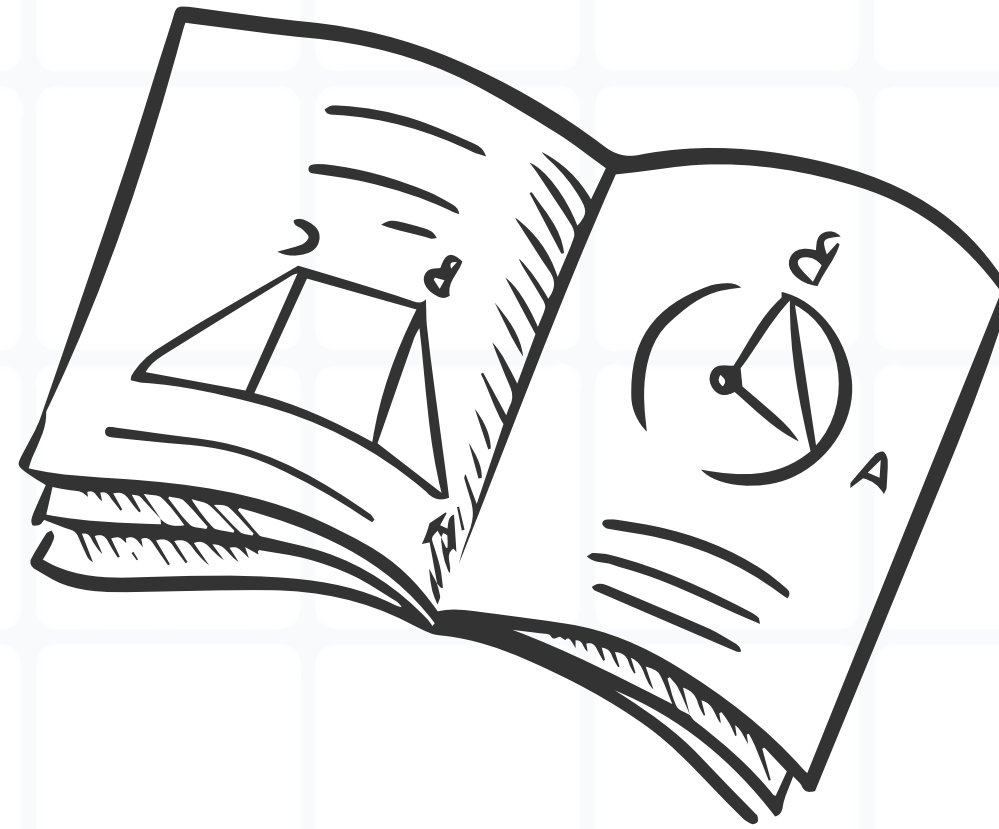


PLAKSHA

HAPPINESS INDEX

- 01.** Conduct Sentiment analysis on Student WhatsApp group chats to determine month-wise happiness levels from 2021-2023.
- 02.** Identify and collect data on features that could potentially affect student happiness levels.
- 03.** Understand which features affect happiness using Machine Learning techniques.
- 04.** Suggest policy implications and recommendations to Plaksha faculty & executives to improve campus happiness :)

LITERATURE REVIEW



Current State of HINGLISH text Sentiment Analysis

- Varsha Thakur, Roshani Sahu, Somya Omer

AIM/METHODOLOGY

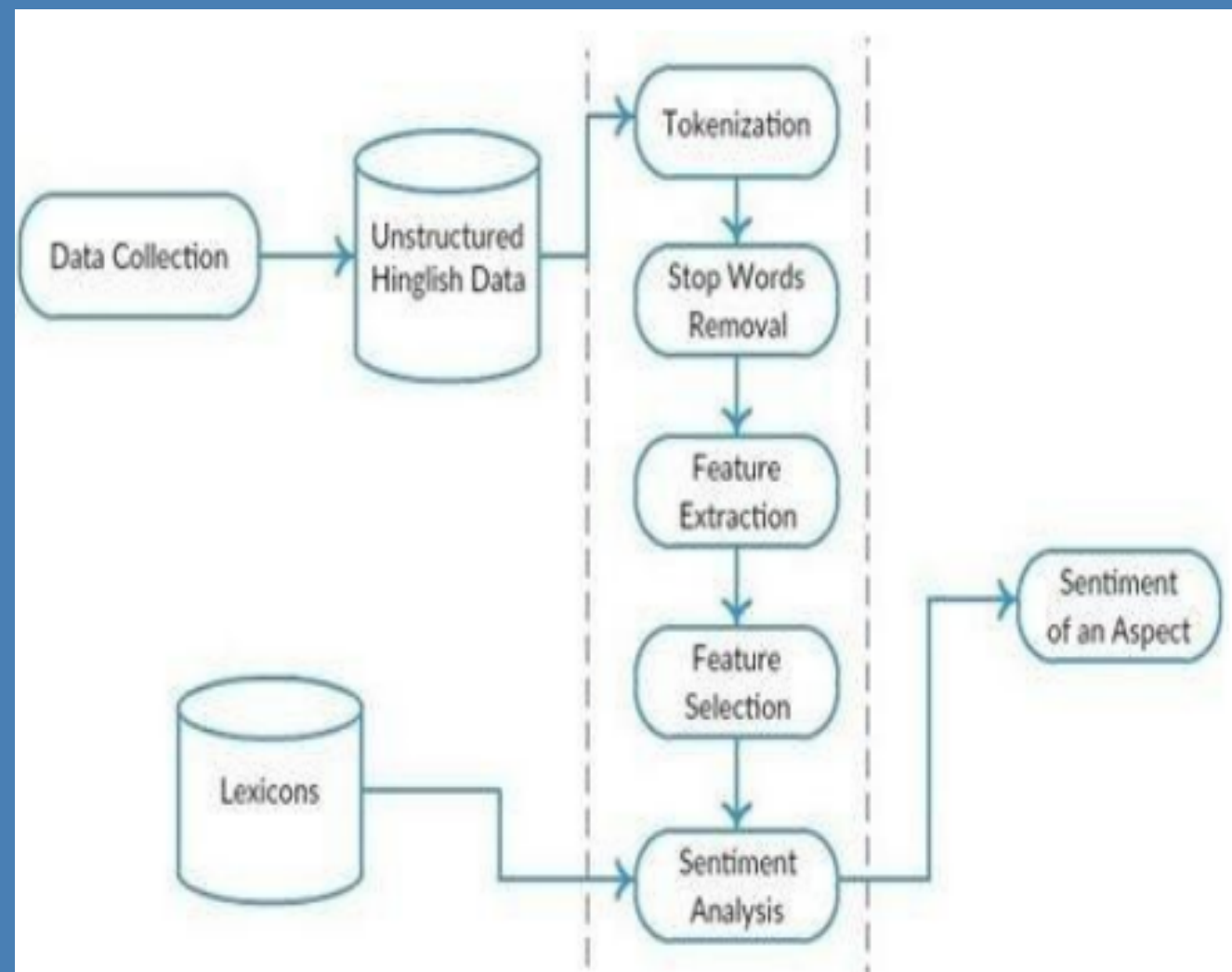
- Reviewed the recent developments in sentiment analysis from both English & Hinglish text
- Presented different problems related to sentiment analysis of Hinglish text, and to explored various classification models and feature extraction techniques used in this area.
- Analyzed the validity of various classification procedures .

INFERENCES

S.no	Author(s)	Language	Approach	Domain	Classifier	Accuracy
4	Kumar Ravi et al.	Hinglish	Machine Learning	News	SentiWordnet	84.67%
-	Tadesse M.M.	Hinglish	Machine Learning	Social Media	NLP-SVM	91%
6	Hassonah M.A. et al.	Hinglish	Machine Learning	Twitter	SVM Hybrid	96.8%
7	Sharma S. et al.	Hinglish	Lexicon Based	Facebook	SentiWordNET	85%

Current State of HINGLISH text Sentiment Analysis

- Varsha Thakur, Roshani Sahu, Somya Omer



INFERENCES

- Text Processing: This stage includes several sub-steps:
 - Tokenization: Breaking down sentences into individual words, phrases
 - Stop Words Removal: Eliminating useless words
 - Feature Extraction: Transforming text into a representation that makes the important textual content available for evaluation. Common methods include unigrams, bigrams, n-grams, term frequency, TF-IDF, POS tagging, and handling negation.
 - Feature Selection: Automatically or manually selecting features that contribute to the predictor of interest, such as variable selection attributes

Utilizing Machine Learning to Predict Happiness Index

- Lexin You

AIM/METHODOLOGY

- Analyzed what factors are responsible for happiness and used this to predict the happiness level of others.
- Collected survey data of 10,000 Chinese adults on 140 features.
- ML Models used: Linear Models, Decision Trees, Random Forest, Gradient Boosting (bagging & boosting).

TABLE VII. THE ERROR TABLE

	Decision Tree	Random Forest	Gradient Boosting
Mean Square Error	0.935	0.609	0.621
Mean Absolute Error	0.672	0.549	0.557

INFERENCES

- Decision Trees were ineffective
- Linear Models were useful after applying Lasso regression in fitting results
- Random Forest & Gradient Boosting models had good accuracy and high R^2 value and were able to rank the features

LIMITATIONS

- Decision Trees were ineffective
- Linear Models were useful after applying Lasso regression in fitting results
- Random Forest & Gradient Boosting models had good accuracy and high R^2 value and were able to rank the features

PAATHSHAALA: Platform for Assessing and Advancing the Happiness Index of Schools and Assisting Learner's Ambitions

- Tushar Sarkara , Disha Shaha

AIM/METHODOLOGY

- Created a model by formulating a parameterized mathematical evaluation criteria that calculates the HI of students and ranks schools accordingly in India.
- Data: Surveyed students, teachers & govt officials, questions related to course level, cleanliness, activities, mental health & parent feedback.
- Sentiment analysis using Latent Dirichlet Allocation (LDA)/Bert on subjective feedback.

LIMITATIONS

- Factors were assigned different weights based on empirical knowledge which could cause bias.
- The study has been done for school students of varied age groups and all features might not be relevant for university students.

INFERENCE

- Subjective and objective questions can be asked to students as well as faculty and executives to assess which features are important.
- Feature Identification: We can use features like exam load, health etc. used in this study

A Study of the Happiness of Chinese University Students and Its Influencing Factors – A Case Study of Beijing Universities. Liang, Chenchen & Sun, Jingdong

A Study on Happiness and Related Factors Among Indian College Students.
Chakraborty, B., Maji, S., Sen, A. et al.

Analysis of World Happiness Report Dataset Using Machine Learning Approaches.
Moaiad Ahmad Khder, Mohammad Adnan Sayfi

An Empirical Study on Sentiment Analysis of Student Happiness Index via Twitter.
Zalikha Zulkifli, Mohamed Imran Mohamed Ariff

DATA

- We collected data from various Plaksha teams like Student Life, Admin, Office of Research, Office of Academics.
- We made sure that we had all the prerequisite permissions.
- We ensured that all the data was anonymised.
- Some data was collected from open-source online resources.
- Most of the data had to be formatted according to our specific needs once we received it in raw form.

DATA

01. Data Collection/Preprocessing

- Monthly Data from December 2021 - October 2023.
- WhatsApp chat data was collected.
- Dummy Variables show presence or absence for the majority of that month
- SL events, holidays, sports are number of occurrences in that month
- library & healthcare show fraction of students issuing books and falling ill.
- All numeric values are scaled

Dummy Variables

- Lounge
- Lockdown
- Music room
- Dance room
- Gym
- Exams

Numeric Data

- SL events
- Holidays
- Sports
- AQI
- Weather
- Humidity
- Campus Strength
- Gender ratio
- Library
- Healthcare
- Maintenance
- Mess

> 20 Features Collected

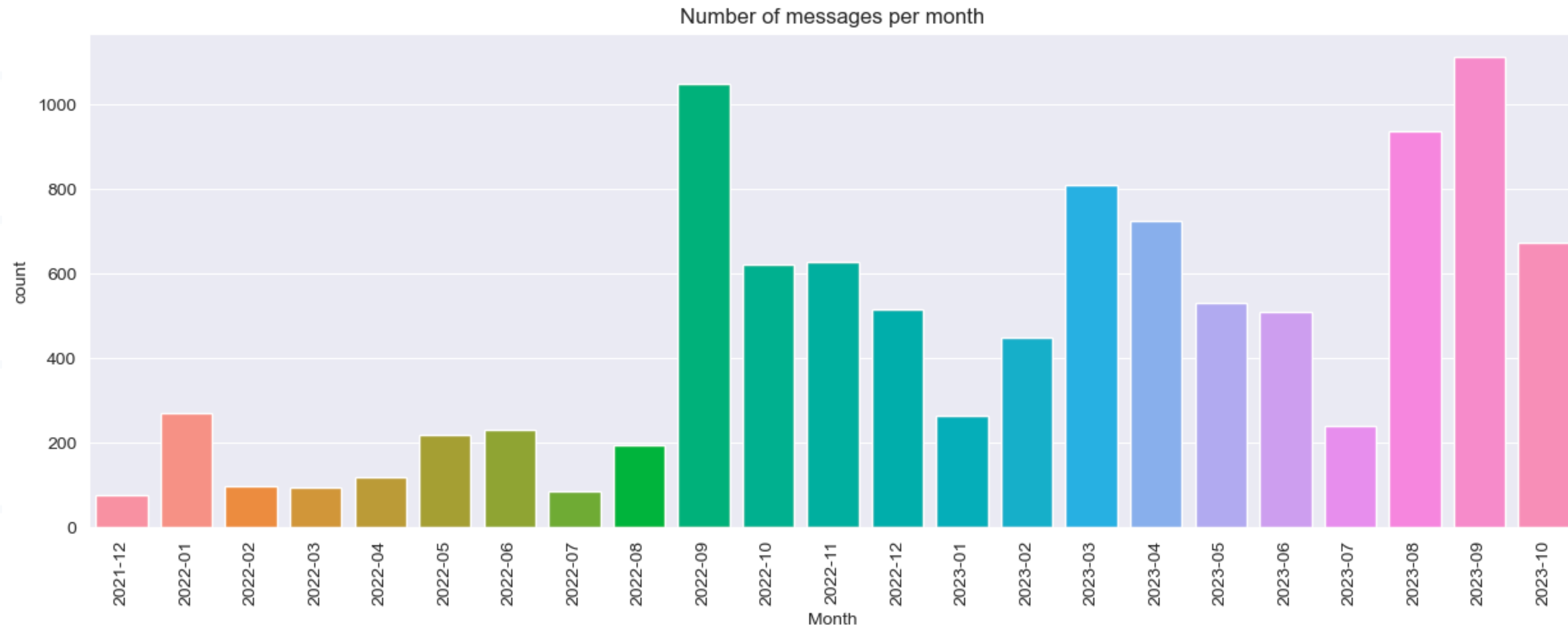
DATA

02. What the data looks like: Features

Months	Year	Lounge	Lockdown	SL events	sports	music room	dance room	gym	exams	holidays	pm25	tempmax	tempmin	temp	precip	no. of students	no. of girls	gender ratio	library abs	healthcare abs	lib_norm	healthcare_norm	sentiment_scores
December	2021	0	0	2	0	0	0	0	1	7	135.677	20.9774	7.96129	13.4194	0.05203	86	29	0.337209302	173	0	2.011627907	0	0.23
January	2022	0	1	0	0	0	0	0	0	31	120	17.7936	8.85484	12.3677	6.20845	86	29	0.337209302	40	88	0.465116279	1.023255814	0.68
February	2022	0	1	0	0	0	0	0	1	28	116.5	21.1286	9.81071	14.6286	1.00139	86	29	0.337209302	106	48	1.23255814	0.558139535	0.48
March	2022	1	0	4	0	0	0	0	0	1	125.871	32.7581	17.2903	24.1452	0.03226	86	29	0.337209302	357	130	4.151162791	1.511627907	0.39
April	2022	1	0	0	0	0	0	0	1	9	123.633	40.0933	22.4267	31.0833	0.0064	86	29	0.337209302	240	130	2.790697674	1.511627907	0.73
May	2022	1	0	0	1	0	0	0	0	3	131.452	40.3	25.7742	32.7936	1.23971	86	29	0.337209302	310	141	3.604651163	1.639534884	0.15
June	2022	1	0	0	1	0	0	0	0	5	119	41.0933	27.1267	33.94	3.7364	86	29	0.337209302	266	169	3.093023256	1.965116279	0.97
July	2022	1	0	0	1	0	0	0	0	31	67.129	35.5419	26.4226	30.5	13.7496	86	29	0.337209302	104	134	1.209302326	1.558139535	0.59
August	2022	1	0	5	1	0	0	0	0	14	70.9333	33.3387	25.1871	28.8903	4.03139	182	56	0.307692308	322	200	1.769230769	1.098901099	0.07
September	2022	1	0	6	6	0	0	0	0	0	79.5	32.07	23.3633	27.33	7.67367	182	56	0.307692308	222	255	1.21978022	1.401098901	0.45
October	2022	1	0	6	1	0	0	0	1	4	101.167	30.3613	18.9387	23.629	1.00761	182	56	0.307692308	237	184	1.302197802	1.010989011	0.81
November	2022	1	0	6	6	0	0	0	0	1	141.433	26.5733	14.02	19.5167	0.0301	182	56	0.307692308	234	230	1.285714286	1.263736264	0.95
December	2022	1	0	5	4	0	0	0	1	0	144.387	22.5258	9.39032	14.7323	0.003	182	56	0.307692308	285	216	1.565934066	1.186813187	0.22
January	2023	1	0	4	2	0	0	0	0	16	184.172	19.7742	8.02258	12.5936	0.94032	182	56	0.307692308	311	287	1.708791209	1.576923077	0.64
February	2023	1	0	2	2	0	0	0	0	0	160.75	26.3786	12.3964	18.3071	0.06786	182	56	0.307692308	195	282	1.071428571	1.549450549	0.12
March	2023	1	0	4	3	0	0	0	1	2	145.71	29.5903	16.029	21.8936	2.561	182	56	0.307692308	143	332	0.785714286	1.824175824	0.51
April	2023	1	0	3	3	0	0	1	0	3	116.852	33.89	19.62	26.3667	0.61133	182	56	0.307692308	165	288	0.906593407	1.582417582	0.04
May	2023	1	0	0	2	1	1	1	1	8	118.931	36.1484	21.9968	29.1774	4.63687	182	56	0.307692308	66	237	0.362637363	1.302197802	0.88
June	2023	0	0	0	0	1	1	1	0	30	96.3	37.49	25.7267	31.3867	4.05357	182	56	0.307692308	47	362	0.258241758	1.989010989	0.33
July	2023	0	0	0	0	1	1	1	0	29	84.1613	33.771	25.8129	29.4613	21.9491	182	56	0.307692308	270	198	1.483516484	1.087912088	0.21
August	2023	0	0	7	5	1	1	1	0	13	115.065	33.8258	25.5226	29.2742	8.0389	321	97	0.302180685	244	467	0.760124611	1.45482866	0.76
September	2023	0	0	3	3	1	1	1	0	1	118.4	33.48	24.25	28.3433	2.27407	321	97	0.302180685	300	400	0.934579439	1.246105919	0.85
October	2023	0	0	2	2	1	1	1	1	2	137.9	31.0774	18.5839	24.1452	0.96	321	97	0.302180685	320	480	0.996884735	1.495327103	0.57

DATA

02. What the data looks like: WhatsApp



DATA

04. Survey

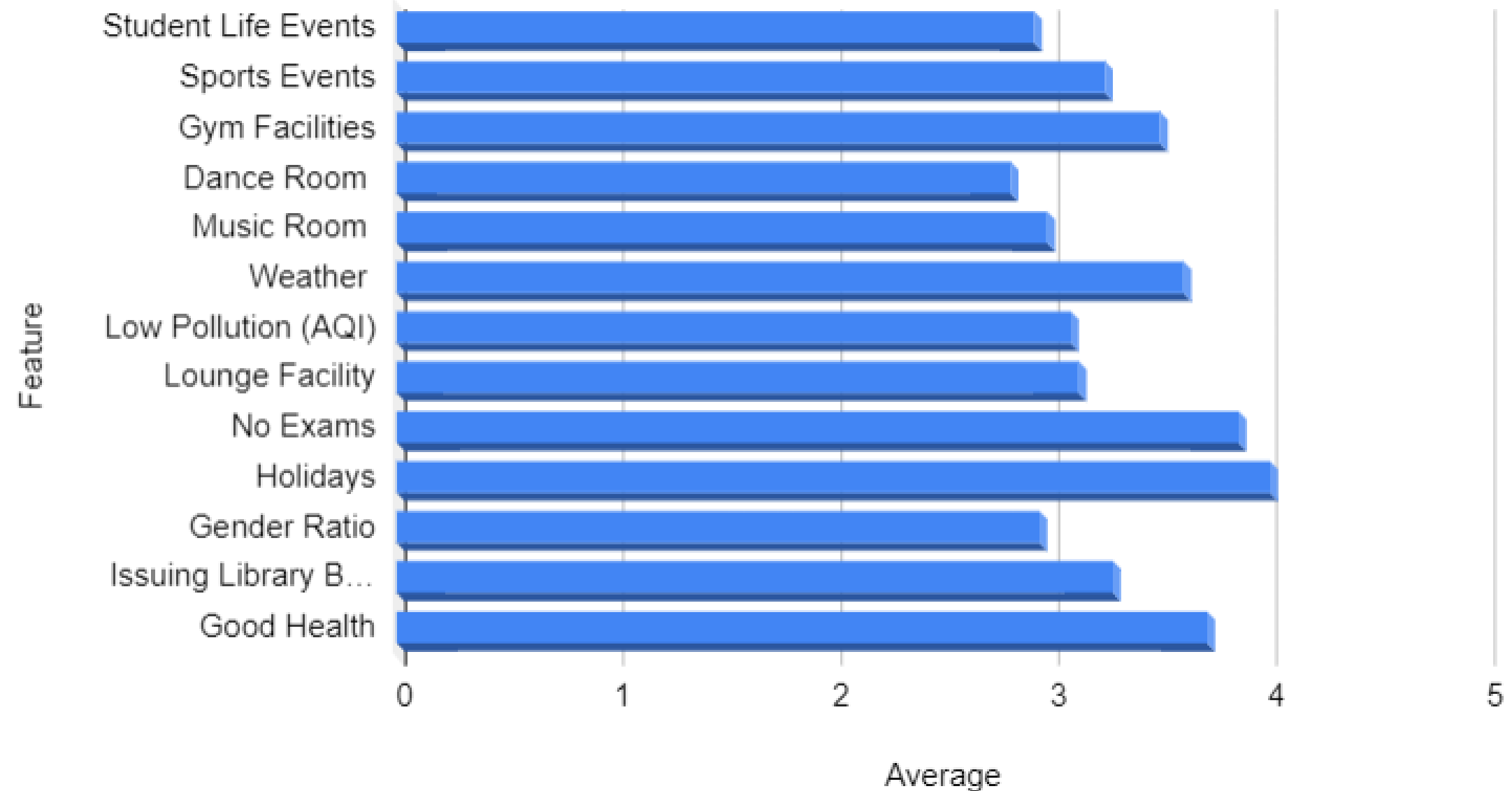
- Establishing student perception
- Ground Truth
- 93 responses
- Main Features for Happiness
 - **Holidays (4.01)**
 - **No Exams (3.86)**
 - **Good Health (3.72)**
 - **Favourable Weather (3.61)**

What makes you Happy at Plaksha :)

1: Least important for happiness
5: Most important for happiness

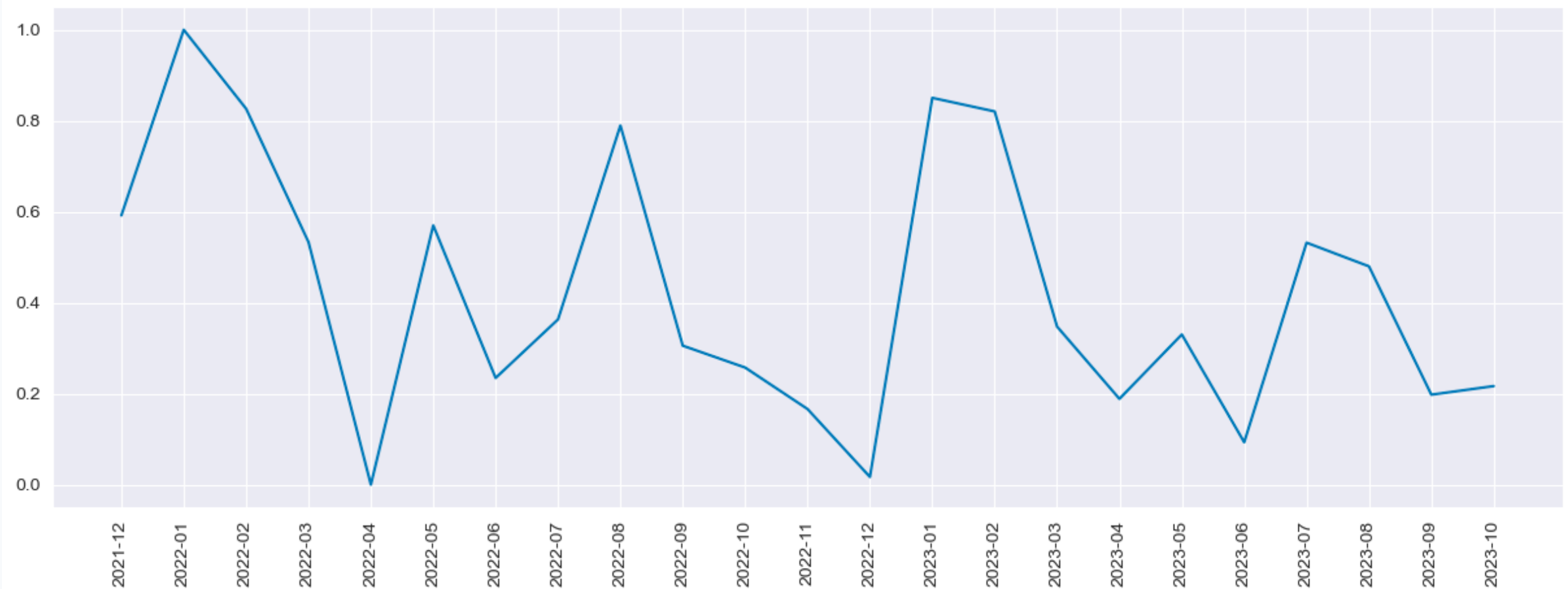
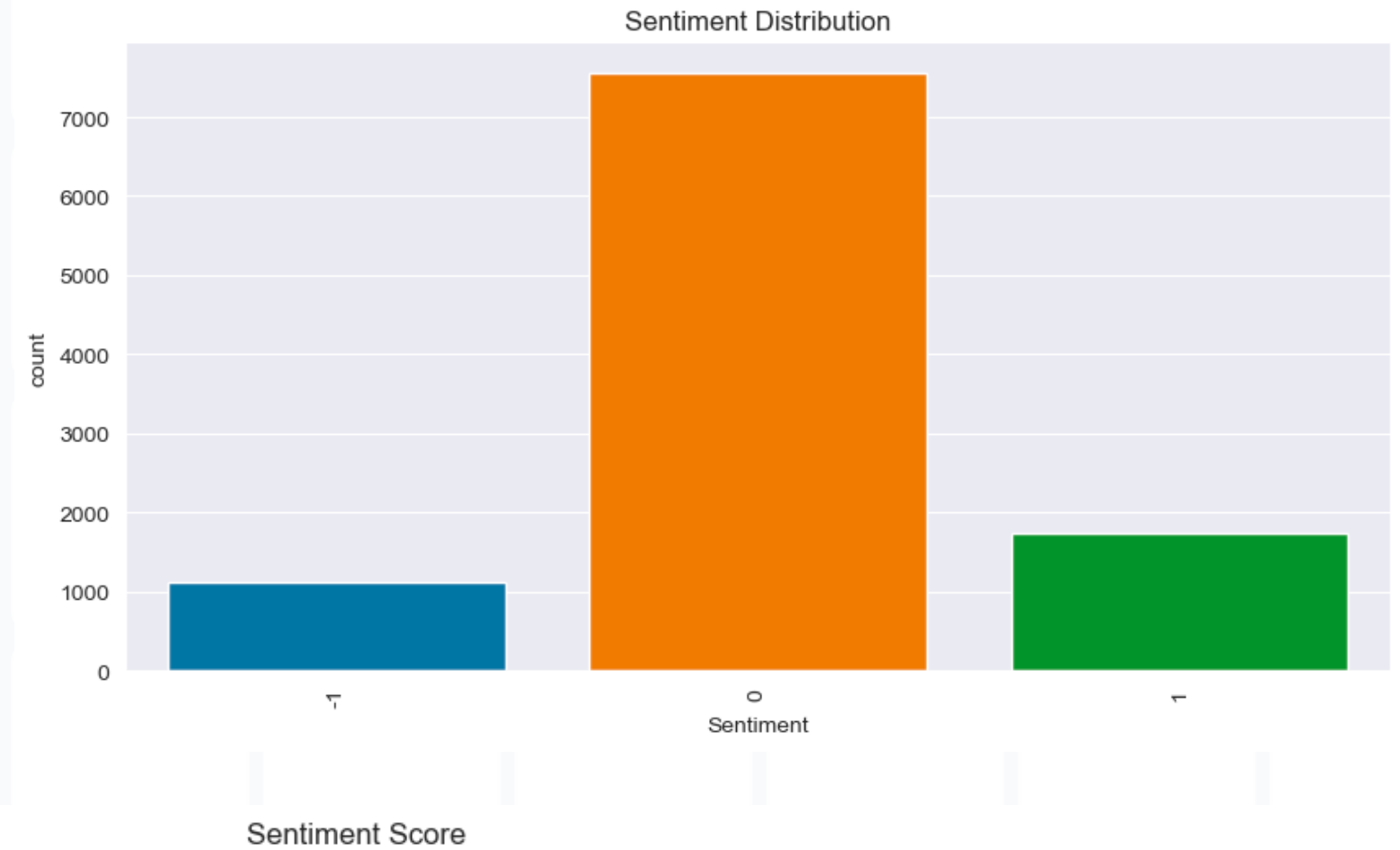


Average vs. Feature



ML METHODOLOGY

- Performed Sentiment Analysis on Plaksha Bulletin WhatsApp chat.
- Anonymised Senders.
- Removed irrelevant messages that could potentially bias results.
- Removed english and hinglish stopwords.
- Used a pre-trained model that uses Bert to perform hinglish sentiment analysis.
- Computed a sentiment score for each month.



ML METHODOLOGY

MODELS USED

- **Random Forests (Bagging):**
 - Provide high predictive accuracy by reducing overfitting.
 - Handle both categorical and continuous data.
 - Automatically select important features.
- **Gradient Boosting (Boosting):**
 - High Predictive Accuracy
 - Sequential Learning: It focuses on correcting errors made by previous models
 - Feature Importance: It can provide feature importance scores
- **XGBoost:**
 - It is a specialized and highly optimized implementation of gradient boosting that is known for its speed, efficiency, and strong performance.

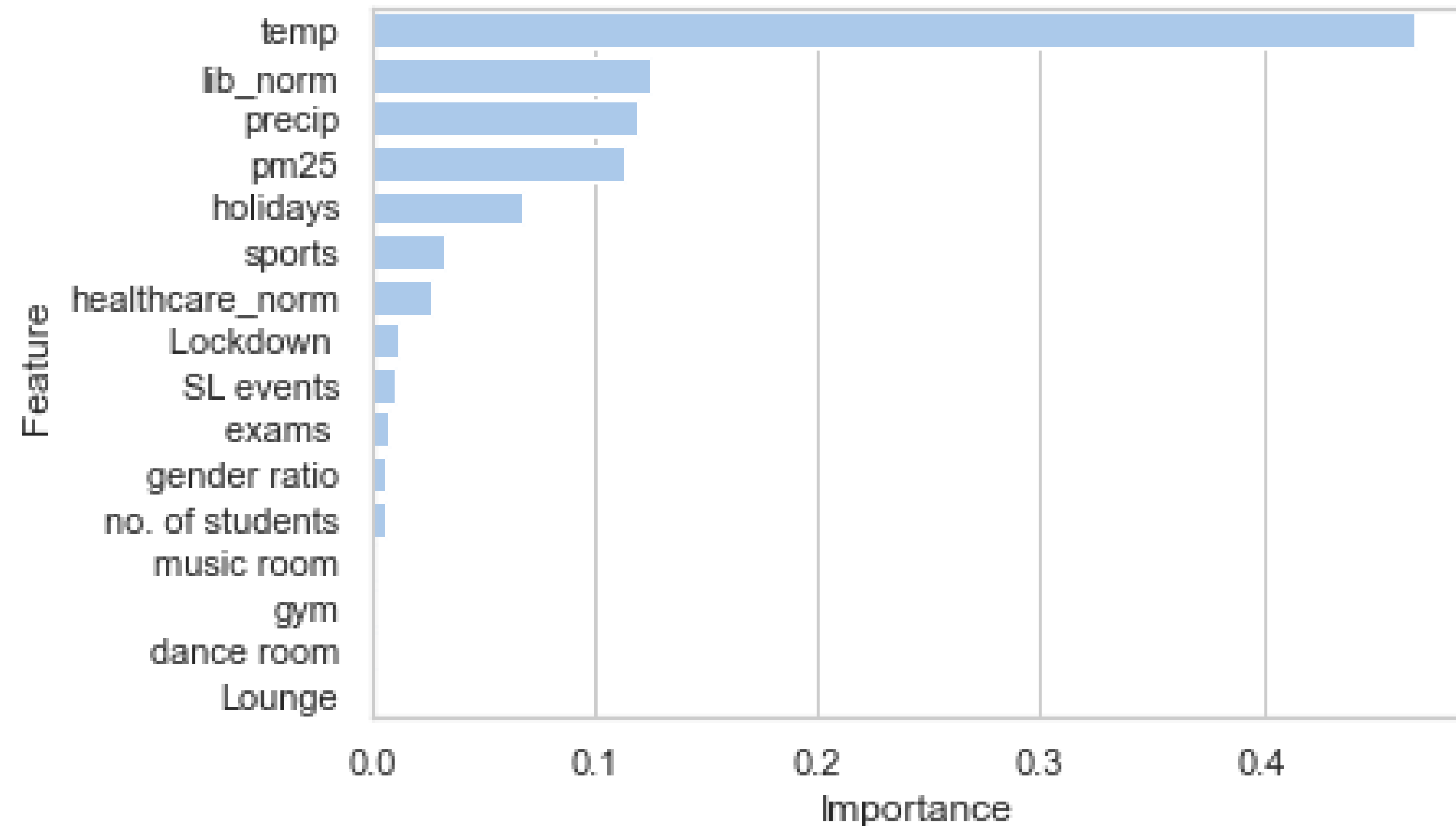
RESULTS

Random Forest

Most Important Features:

- Favourable Weather : 0.467387
- Issuing Library books: 0.124681
- Holidays: 0.067196
- Sports Activities: 0.032735

Feature Importance



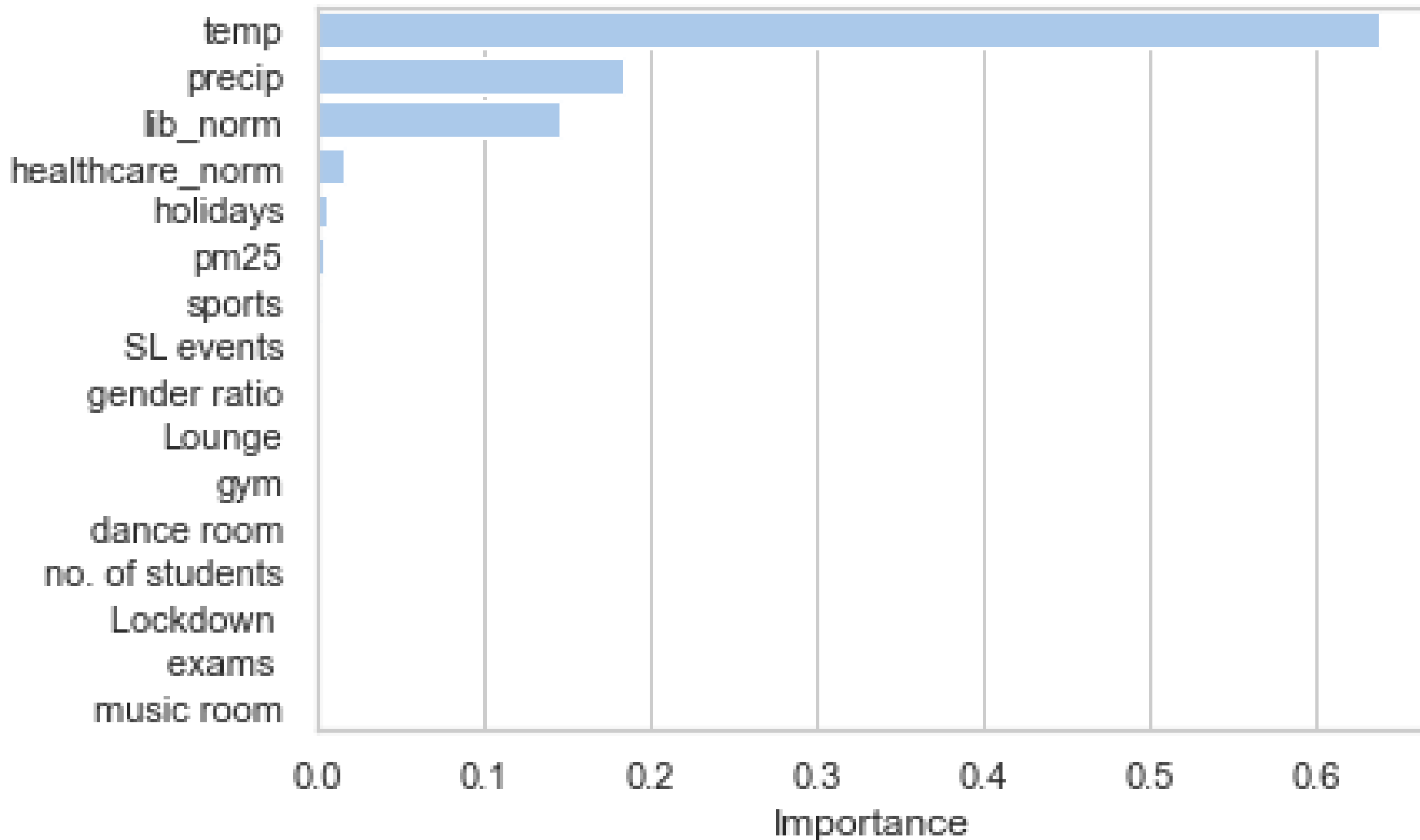
RESULTS

Gradient Boosting

Most Important Features:

- Favourable Weather : $6.367133e-01$
- Issuing Library books: $1.471076e-01$
- Good Health: $1.564145e-02$
- Holidays: $6.639693e-03$

Gradient Boosting Feature Importance



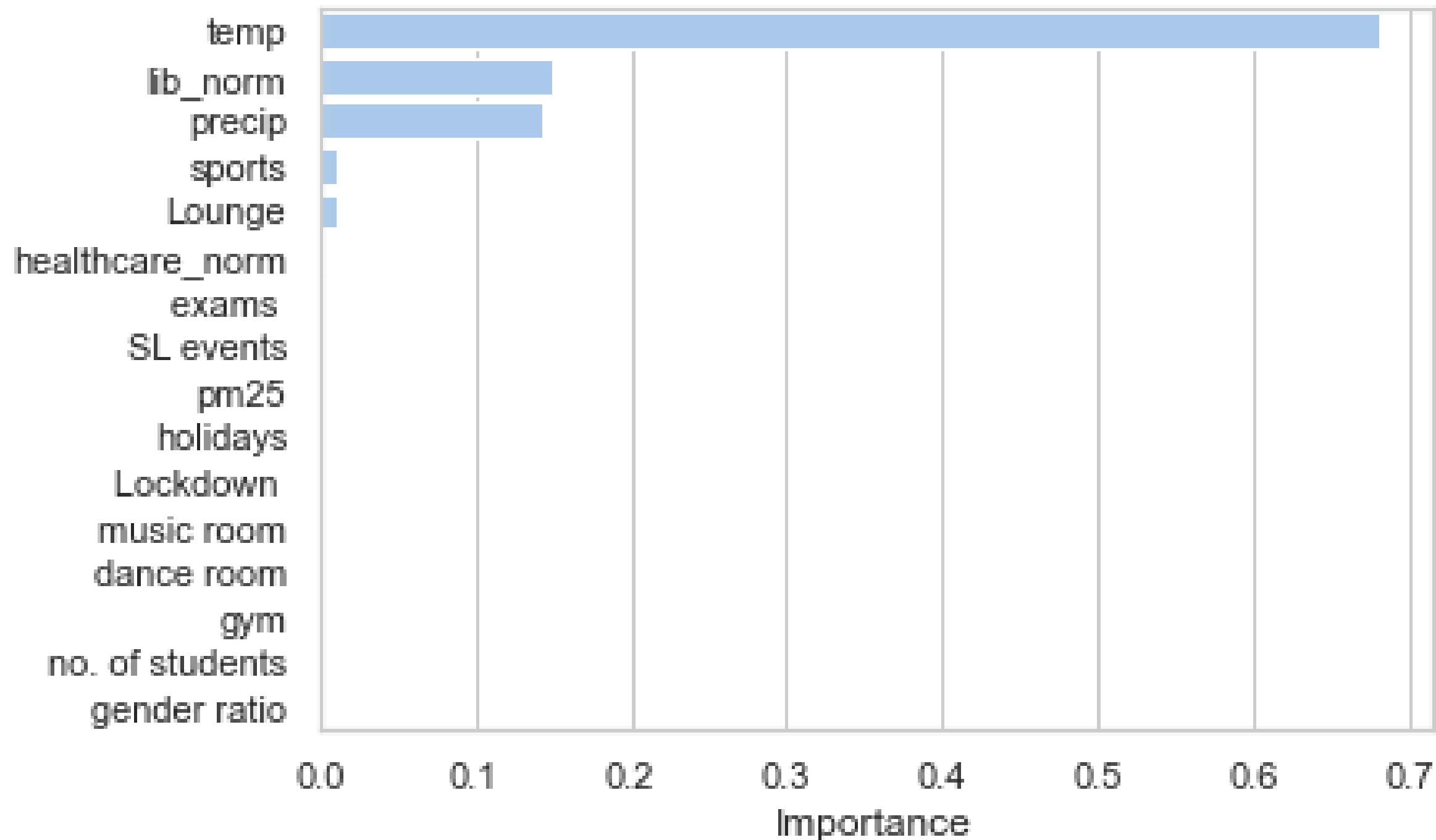
RESULTS

XGBoosting

Most Important Features:

- Favourable Weather : 0.679726
- Issuing Library books: 0.149499
- Sports events: 0.011939
- Lounge: 0.011573

XGBoost Feature Importance



RESULTS

Most Important Features from MODEL

- Favourable Weather
- Issuing Library Books
- Good Health
- Sports Events
- Holidays

Most Important Features from SURVEY

- Holidays (4.01)
- No Exams (3.86)
- Good Health (3.72)
- Favourable Weather (3.61)

	Model	MAE	MSE	RMSE
0	Random Forest	0.255403	0.109871	0.331468
1	Gradient Boosting	0.302222	0.139778	0.373869
2	XGBoost	0.288428	0.137628	0.370983

Similar values for Mean Absolute Error, Mean Squared Error & Root Mean Squared Error for all 3 models

IMPLICATIONS

- **Campus Infrastructure and Facilities:** Consider investing in climate-controlled facilities or shaded areas to mitigate the impact of adverse weather conditions and create comfortable study and leisure spaces for students.
- **Library Services:** Ensure an efficient and well-stocked library with a seamless book issuing process to support students' academic pursuits and research.
- **Health and Wellness Services:** Prioritize student health and wellness programs, including mental health services, access to healthcare, and fitness facilities to promote good health among students and staff.
- **Sports and Recreational Facilities:** Develop and maintain sports infrastructure, organize regular sports events, and encourage participation in physical activities to promote a healthy and active lifestyle.
- **Academic Calendar Planning:** Consider the impact of holidays and academic schedules on students' well-being and academic performance. Balance the academic workload and provide sufficient breaks.
- **Student Engagement:** Create opportunities for students to engage in extracurricular activities, including sports, cultural events, and clubs, to enhance their overall satisfaction and well-being.

CHALLENGES/LIMITATIONS

- **Limited and Incomplete Data:** Plaksha data spans only two years, which may not provide a comprehensive understanding of long-term trends or patterns. Some data is incomplete, which can affect the accuracy.
- **Use of WhatsApp Chats:** WhatsApp messages can be informal, context-dependent, and contain slangs, or abbreviations that make sentiment analysis challenging. Exploring alternative data sources, such as surveys, interviews, or structured feedback forms, could provide more reliable sentiment data.
- **Missing Data on Important Factors:** The absence of crucial data elements, such as mess data (information related to food services) and complaints data, limits the comprehensiveness of the analysis.
- **Inability to Perform Neural Networks:** Due to the limited dataset, it may not be feasible to implement more complex machine learning models, such as Neural Networks
- **Bias in Data due to Changing Student Population:** Over time, the student population at Plaksha has changed with new cohorts of students joining. This can lead to bias in the data, as the experiences and sentiments of newer students may differ from those of earlier cohorts.



THANK YOU

