

The background features a complex network of thin, light gray lines and dots, primarily concentrated on the left side, resembling a neural network or data structure. Scattered across the entire background are various thin-lined triangles of different sizes and orientations. Some triangles are solid light gray, while others are just outlines. The overall aesthetic is clean, modern, and tech-oriented.

# **Personality Prediction System**

---

# Introduction

In this work, we intent to develop a system that can predict one's personality based on the post or status one uploads on the social media network.

This study is based on the prediction of Big Five Personality, MBTI (Myers-Briggs Type Indicator)

- I-E: Extraversion (E) or Introversion (I)
- N-S: Sensing (S) or INtuition (N)
- T-F: Thinking (T) or Feeling (F)
- J-P: Judging (J) or Perceiving (P)



# Why Personality Prediction ?



- Recruitment systems
- Personal Counseling Systems
- Computational Advertising,
- Marketing Science
- Enhanced Human-Computer Interaction
- Bank Credit Scoring Systems.

People often update statuses, posts, comments to express their feelings and opinions on social media. These expressions can be used to characterize the individual's behavior, personality, and characteristics of their thought patterns.

**NF***Valuing**Manifesting universal values  
and valuing people***Possible****NT***Visioning**Pulling people with ideas  
to an optimistic future*

<b>ENFJ</b> <b>Teacher</b> Smooth talking charmers. Very inspiring & motivational. Often clergy. People leaders & persuaders. Great salespeople. Very relationship-oriented. Like to motivate groups.	<b>INFJ</b> <b>Counselor</b> Work is to inspire others to achieve great things. Great visionaries of human possibilities. Serious academicians. Often professors or offer themselves to a religious order.	<b>INTJ</b> <b>Mastermind</b> If they say they are going to do something, they do it. Likely to be corporate leaders, scientists. Believe everything has room for improvement. Superior planners and visionaries of systems.	<b>ENTJ</b> <b>Field Marshall</b> Very leadership-oriented. Likely to be top executives, business persons. Big on reducing inefficiency, ineffectiveness. Take charge people. Can be overwhelming to less outgoing types.
<b>ENFP</b> <b>Champion</b> Second only to ESFPs for fun. Want lives filled with excitement and romance. Very enthusiastic and creative. Often teachers, artists, writers. Great need for diversity and change.	<b>INFP</b> <b>Healer</b> Noble servants aiding society. Different from ISFPs, they try to tackle long term problems. Often psychologists or counselors. Want to save the whales and rainforests.	<b>INTP</b> <b>Architect</b> Deepest analysts of problems to be solved. Often physicists, scientists. Most stoic of types. Critical thinkers.	<b>ENTP</b> <b>Inventor</b> Want one exciting challenge after another. Love to problem solve. Good at analysis, consider themselves full of ingenuity and ideas. Often involved in computer, systems analysis, design.
<b>ESFP</b> <b>Performer</b> Number one in fun and enthusiasm. Always invite ESFPs to your party. The most generous of all types. Warm, friendly, vibrant people. Excellent at customer service.	<b>ISFP</b> <b>Composer</b> Quietly harmonious with world. Very observing, benevolent. Inclined toward work with people in need. Work to solve problems of the immediate such as homeless, stopping hunger.	<b>ISTP</b> <b>Operator</b> Ready to try anything once. Pushed with the rush of life. Seek excitement. A love of tools and the utility they offer. Inclined toward mechanical devices, can take apart & reassemble anything.	<b>ESTP</b> <b>Promoter</b> Excitement seekers. Never feel more alive than when taking risks. Great negotiators on the front end. Excellent promotional & entrepreneurial capabilities if someone else follows through.
<b>ESFJ</b> <b>Provider</b> Hosts & hostesses. Graciousness of this type makes them excellent at entertaining, coordinating. May be teachers, nurses. Very conscious of appearances, should/shouldn'ts.	<b>ISFJ</b> <b>Protector</b> A high sense of duty. Upholders of family tradition. Often found in traditional helping professions including nursing, elementary education, etc.	<b>ISTJ</b> <b>Inspector</b> Doers of what should be done. Masters at compiling practical details and adding finishing touches. Get-it-done people. Superb administrators. Duty bound & dogmatic, often military.	<b>ESTJ</b> <b>Supervisor</b> Administrators, workers, pillars of strength in community. Logical, matic, parents, employees. Often promoted to management positions. Dependable, consistent, straightforward.

**Personal****Logical****SF***Relating**Including and building trustworthiness***Present****ST***Directing**Action from a strategic perspective*



# Literature Review

- Comparison of machine learning algorithms for content based personality resolution of tweets

Shruti Garg, Ashwani Garg,

(<https://www.sciencedirect.com/science/article/pii/S2590291121000747>)


- Personality Prediction Using Machine Learning

Shilpa R, Supriya V, Sweta Prasad, Vinaya Varshini R, Uday Shankar SV

(<https://saejournal.com/wp-content/uploads/2021/07/Personality-Prediction-Using-Machine-Learning.pdf> )

- ABOUT MYER-BRIGGS TYPE INDICATOR

<https://rismakov.com/mbti-prediction/category/xgboost>



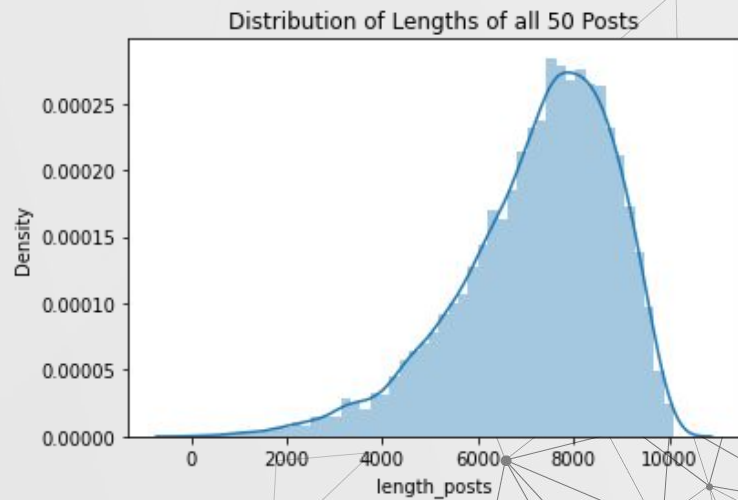
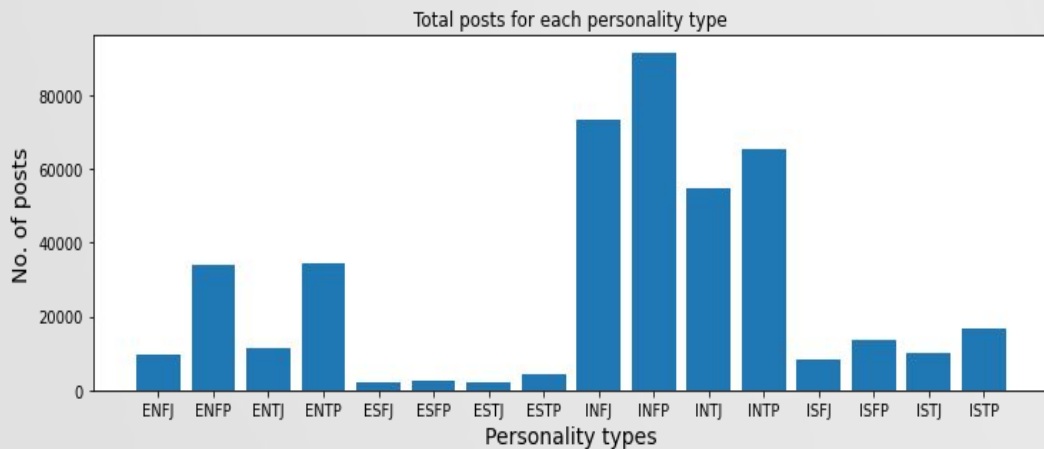
# Dataset

The dataset used in this study is from the Myers–Briggs personality dataset which is openly available on Kaggle (<https://www.kaggle.com/datasnaek/mbti-type> ).

- 8675 unique value
- First column : MBTI Personality Type
- Second Column : 50 things they have posted.

	type	posts
8670	ISFP	'https://www.youtube.com/watch?v=t8edHB_h908  ...
8671	ENFP	'So...if this thread already exists someplace ...
8672	INTP	'So many questions when i do these things. I ...
8673	INFP	'I am very conflicted right now when it comes ...
8674	INFP	'It has been too long since I have been on per...

# Visualizing Dataset





## Tools used for Development

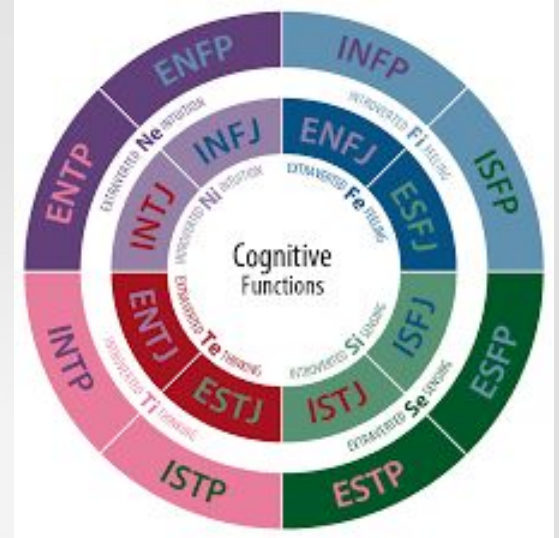
- NLTK (natural language processing toolkit) and XGBoost
- Frequency–Inverse Document Frequency
- Pandas, numpy, re, seaborn, matplotlib and sklearn are other Python libraries are used



# Pipeline

## 1. Preprocessing of Dataset

- a. URL's and Stop words were removed from the dataset.
- b. Splitting the MBTI personality into four letters and binarising it.
- c. the text was lemmatised, i.e., inflected forms of the words were transformed into their root words



## 2. Feature Engineering

- a. Vectorise with Count and Term Frequency–Inverse Document Frequency (TF–IDF)

# Pipeline (Contd.)

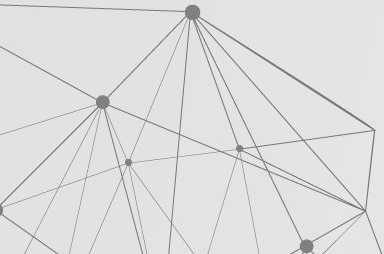
## 3. Splitting the Features:

X: user posts in TF-IDF representation

Y: Personality type in Binarised MBTI form

## 4. Training Model

- a. In total, 70% of the data was used as the training set and 30% of the data was used as the test set. The model was fit onto the training data and the predictions were made for the testing data.



## 5. Evaluating Model

### Random Forest Classifier

IE: Introversion (I) / Extroversion (E) Accuracy: 77.33%

NS: Intuition (N) / Sensing (S) Accuracy: 86.03%

FT: Feeling (F) / Thinking (T) Accuracy: 67.66%

JP: Judging (J) / Perceiving (P) Accuracy: 62.80%

### XGBoost Classifier

IE: Introversion (I) / Extroversion (E) Accuracy: 77.65%

NS: Intuition (N) / Sensing (S) Accuracy: 86.06%

FT: Feeling (F) / Thinking (T) Accuracy: 68.77%

JP: Judging (J) / Perceiving (P) Accuracy: 64.83%

### KNN Classifier

IE: Introversion (I) / Extroversion (E) Accuracy: 76.67%

NS: Intuition (N) / Sensing (S) Accuracy: 85.82%

FT: Feeling (F) / Thinking (T) Accuracy: 54.70%

JP: Judging (J) / Perceiving (P) Accuracy: 40.20%

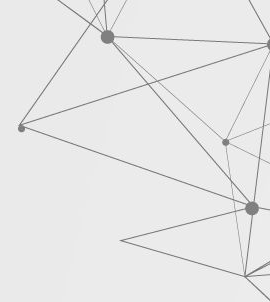




# Phase 1- Result and Conclusion

Out of all the models, seen above we see that on an average XG Boost gives relatively good performance, hence we choose it to build our Personality prediction model. This will be beneficial as XGBoost model can even be used to evaluate and report on the performance on a test set for the model during training.

The results show that the methodology has better accuracy and reliability in comparison to other existing methods. Regarding the knowledge contribution in this paper, the presented methodology significantly improved the accuracy of recognising the Intuition (I)–Sensing (S) and Introversion (I)–Extroversion (E) personality categories, as well as slightly better accuracy for the Judging (J)–Perceiving (P) personality category. This can effectively assist NLP practitioners and psychologists in regards to identification of personality types and associated cognitive processes.



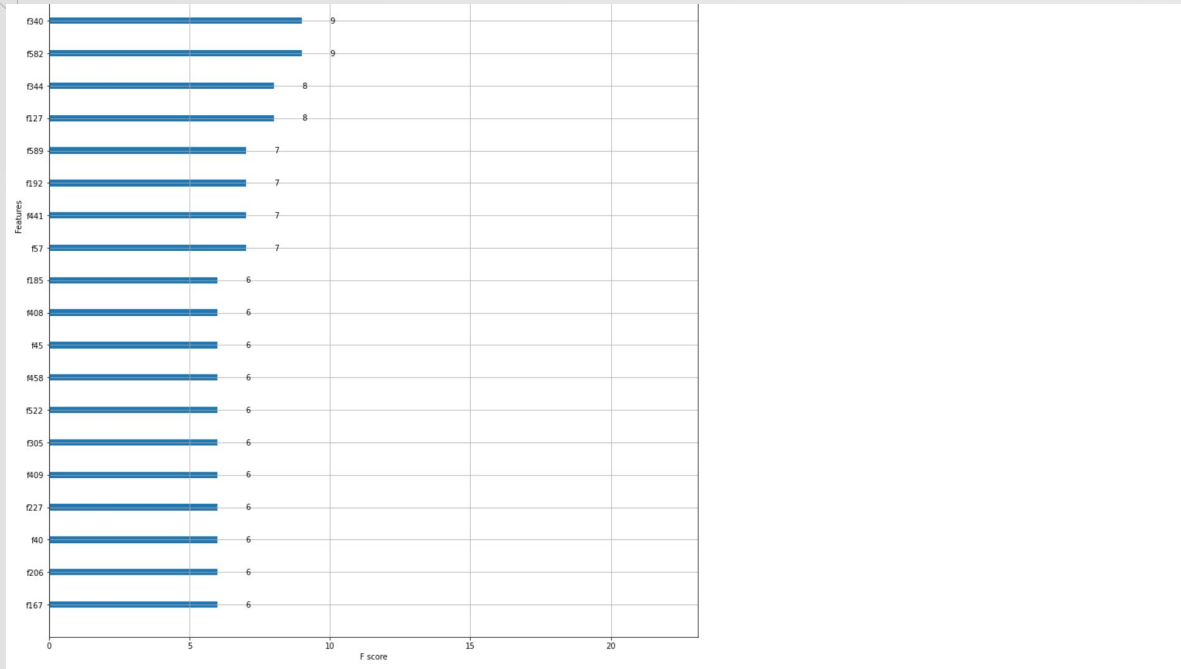
Abstract geometric patterns in the top corners of the slide. The top-left corner features a network of thin grey lines connecting several black dots, forming various triangular shapes. The top-right corner shows a similar but more complex network of lines and dots. The background is a solid light grey.

# **PHASE 2**

## **(The Testing Phase)**

# WHY XGBOOST???

## Feature Importance

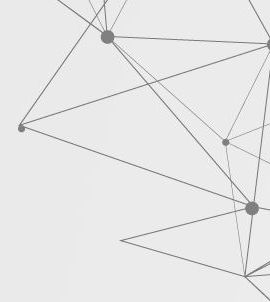


XGBOOST automatically provide estimates of feature importance from a trained predictive model.



# Hyperparameter Tuning

**The parameters to consider tuning are:**

- The number and size of trees (`n_estimators` and `max_depth`).
  - The learning rate and number of trees (`learning_rate` and `n_estimators`).
  - The row subsampling rates (`subsample`)
- 

# Our Tuning FORMULATION!!

```
IE: Introversion (I) / Extroversion (E) ...
* Best: -0.508298 using {'learning_rate': 0.2, 'n_estimators': 200}
* -0.508298 (0.012022) with: {'learning_rate': 0.2, 'n_estimators': 200}
* -0.511190 (0.014109) with: {'learning_rate': 0.2, 'n_estimators': 300}
* -0.515801 (0.013401) with: {'learning_rate': 0.3, 'n_estimators': 200}
* -0.523666 (0.014271) with: {'learning_rate': 0.3, 'n_estimators': 300}
NS: Intuition (N) / Sensing (S) ...
* Best: -0.395608 using {'learning_rate': 0.2, 'n_estimators': 200}
* -0.395608 (0.002532) with: {'learning_rate': 0.2, 'n_estimators': 200}
* -0.407233 (0.004062) with: {'learning_rate': 0.2, 'n_estimators': 300}
* -0.418423 (0.003287) with: {'learning_rate': 0.3, 'n_estimators': 200}
* -0.436628 (0.004779) with: {'learning_rate': 0.3, 'n_estimators': 300}
FT: Feeling (F) / Thinking (T) ...
* Best: -0.559850 using {'learning_rate': 0.2, 'n_estimators': 200}
* -0.559850 (0.019320) with: {'learning_rate': 0.2, 'n_estimators': 200}
* -0.566251 (0.021509) with: {'learning_rate': 0.2, 'n_estimators': 300}
* -0.575489 (0.024986) with: {'learning_rate': 0.3, 'n_estimators': 200}
* -0.588334 (0.027208) with: {'learning_rate': 0.3, 'n_estimators': 300}
JP: Judging (J) / Perceiving (P) ...
* Best: -0.650526 using {'learning_rate': 0.2, 'n_estimators': 200}
* -0.650526 (0.014655) with: {'learning_rate': 0.2, 'n_estimators': 200}
* -0.662368 (0.015849) with: {'learning_rate': 0.2, 'n_estimators': 300}
* -0.667228 (0.012558) with: {'learning_rate': 0.3, 'n_estimators': 200}
* -0.685591 (0.018371) with: {'learning_rate': 0.3, 'n_estimators': 300}
```

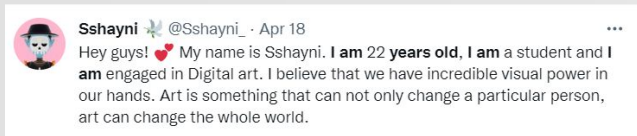
- learning rate as 0.2
- number of estimators as 200
- maximum depth of tree as 2
- number of threads as 8



# TEST CASES

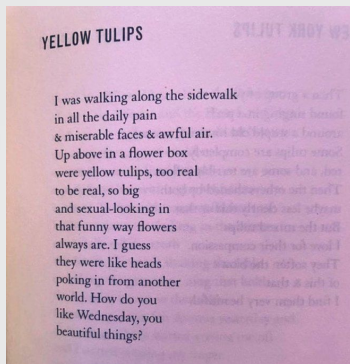
## Let'S PREDICT!!

### Random post



The result is: INFP

### A Poem



The result is: INFP

### A Tweet



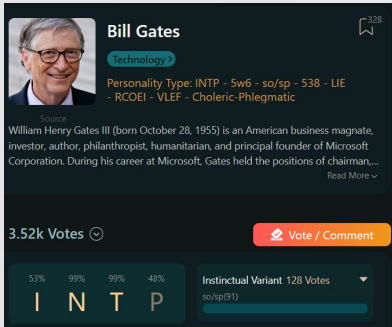
The result is: INFJ

# Failures!!

## Bill Gates tweet's

The result is: INFP

Reality:



**Bill Gates**  
Technology >

Personality Type: INTP - 5w6 - so/sp - 538 - LIE  
- RCOEI - VLEF - Choleric-Phlegmatic

Source  
William Henry Gates III (born October 28, 1955) is an American business magnate, investor, author, philanthropist, humanitarian, and principal founder of Microsoft Corporation. During his career at Microsoft, Gates held the positions of chairman,...

3.52k Votes

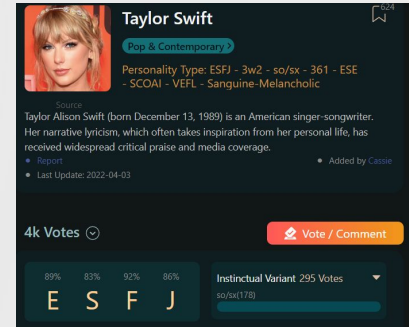
Vote / Comment

53% 99% 99% 48%  
**I N T P**  
Instinctual Variant 128 Votes  
so/sp(91)

## Taylor Swift's tweet's

The result is: INFP

Reality:



**Taylor Swift**  
Pop & Contemporary >

Personality Type: ESFJ - 3w2 - so/sx - 361 - ESE  
- SGOAI - VEFL - Sanguine-Melancholic

Source  
Taylor Alison Swift (born December 13, 1989) is an American singer-songwriter. Her narrative lyricism, which often takes inspiration from her personal life, has received widespread critical praise and media coverage.

- Report
- Last Update: 2022-04-03
- Added by Cassie

4k Votes

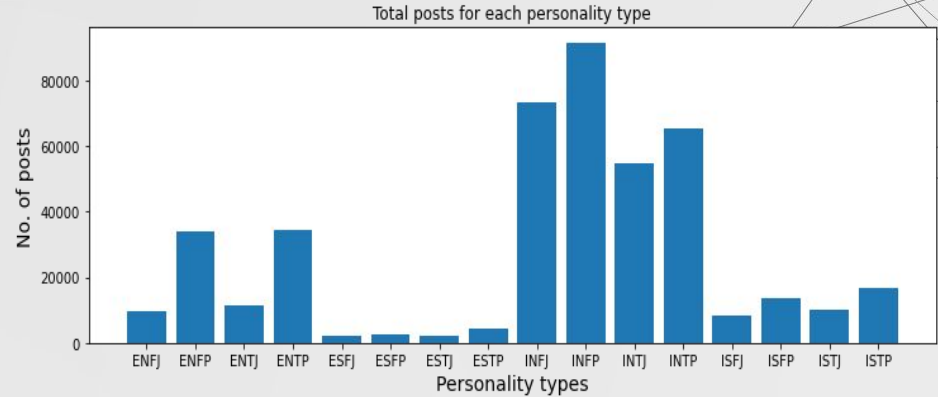
Vote / Comment

89% 83% 92% 86%  
**E S F J**  
Instinctual Variant 295 Votes  
so/sx(178)

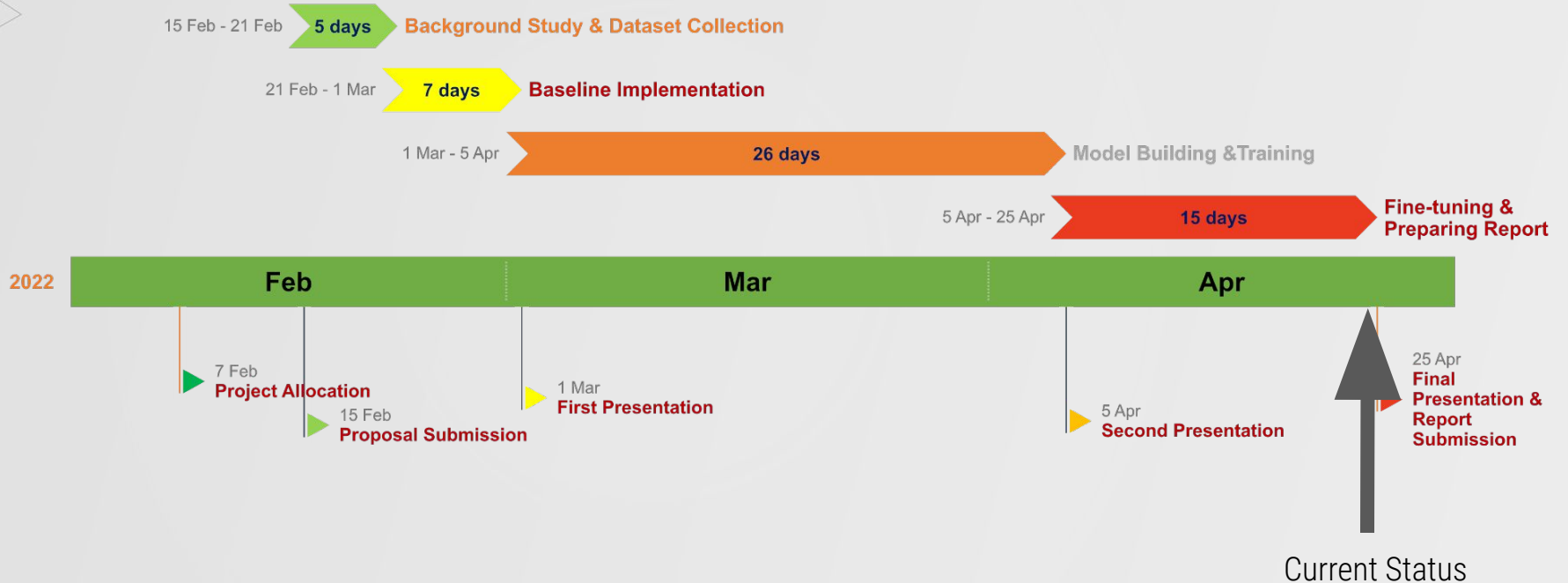
## Reasons for failures

- **Data Imbalance**
- **Myers-Briggs provides inconsistent, inaccurate results**

Implementing with more data, natural language processing methods, or newer text features will help us predict it accurately



# Timeline and Progress





# Work/Code

[https://colab.research.google.com/drive/1kwwCyo\\_-HA9hlXDbXzT7xYVa-Qjg2L1y?usp=sharing](https://colab.research.google.com/drive/1kwwCyo_-HA9hlXDbXzT7xYVa-Qjg2L1y?usp=sharing)



# References

1. <https://www.personality-database.com/profile?pid=1&sort=top>
2. <https://www.sciencedirect.com/science/article/pii/S2590291121000747>
3. <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6839354.pdf>
4. <https://www.kaggle.com/datasets/datasnaek/mbti-type>



# Group Members

1. Darshan Patil 21120006
  2. Shoaib Alam 21250011
  3. Iram Nawab 21250024
  4. G. V. Sai 21210036
- 