

CLUB RECRUTER

USING NLP AND MACHINE LEARNING



VIDEO LINK

[https://drive.google.com/drive/folders/1IGc-HKYCE_L5NgStW3KrGYAhPmJNb0b5?
usp=sharing](https://drive.google.com/drive/folders/1IGc-HKYCE_L5NgStW3KrGYAhPmJNb0b5?usp=sharing)

MEET OUR PROJECT MEMBERS



Surya S

20BCE1071



Nilavan I

20BCE1080



Sumegh S Gonugade

20BCE1100



Nithish Kanna G

20BCE1304



INTRODUCTION

Our NLP project involves predicting the most suitable department for an individual based on their personality traits. By analyzing the language used in their social media posts and correlating it with different personality traits, we aim to identify the most relevant department for a given student. Using a combination of techniques such as text preprocessing, feature extraction, and machine learning, our model is trained on a dataset of personality traits and position descriptions to predict the best fit department. The project has the potential to assist club committees in selecting the most suitable department for a student's personality traits, resulting in a better fit and more suitable position within the club.

METHODOLOGY

Methodology for Club Recruitment using NLP and ML involves the following steps:

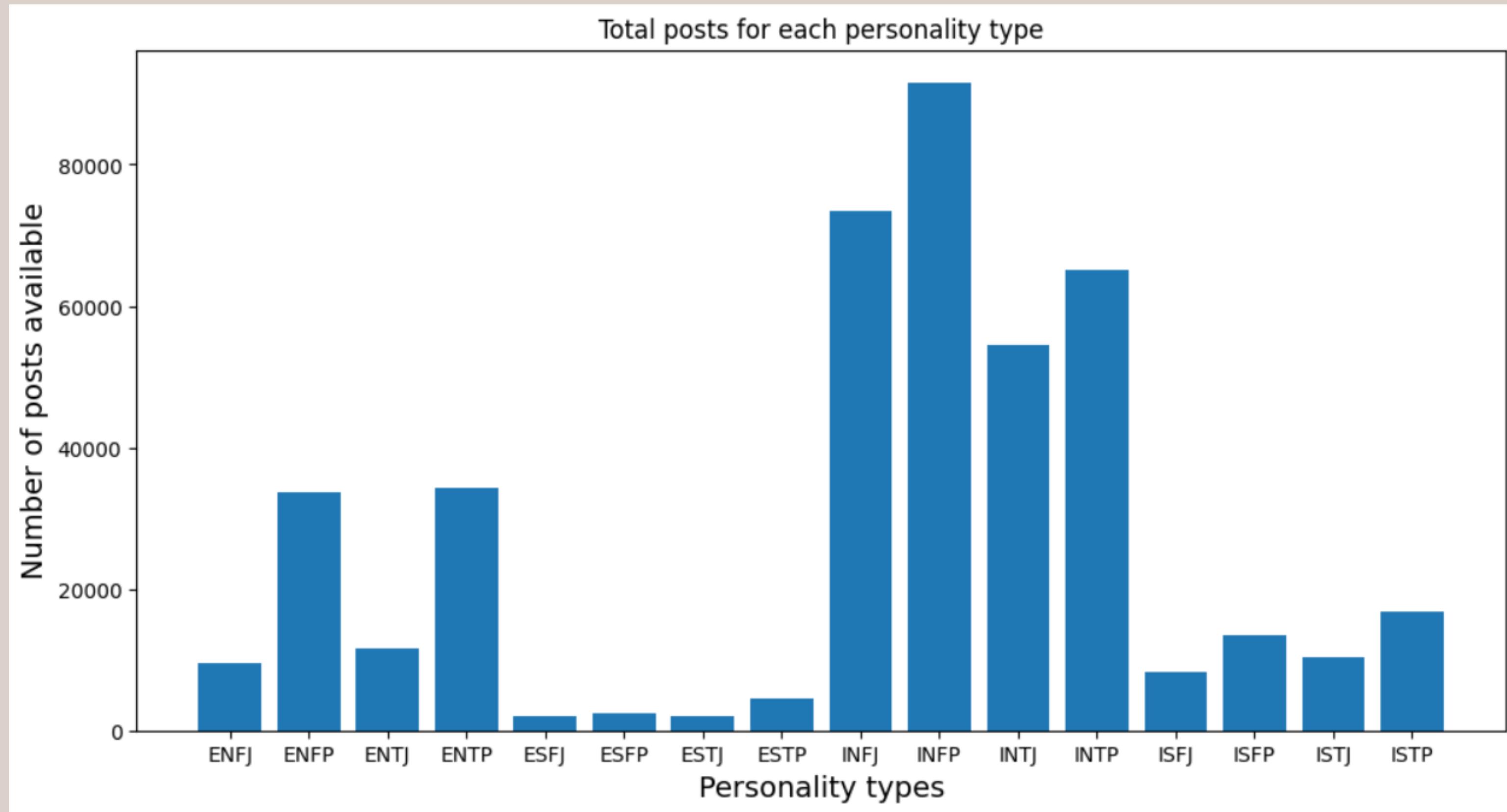
- Data Pre-processing
- Model selection
- Club Department Recruitment

DATA PRE-PROCESSING

This involves extracting of data, cleaning of data and preparing it for the ML models.

- Data Extraction: Data for this work is taken from Kaggle.
- Data Organizing: Dataset is analyzed and posts are split into 16 personalities.
- Data Cleaning: NLP toolkit is used to tokenize the posts using bag_of_words model.
- Data Splitting: Data is split into train and test in 80:20 for the ML models.

Total posts for each personality type

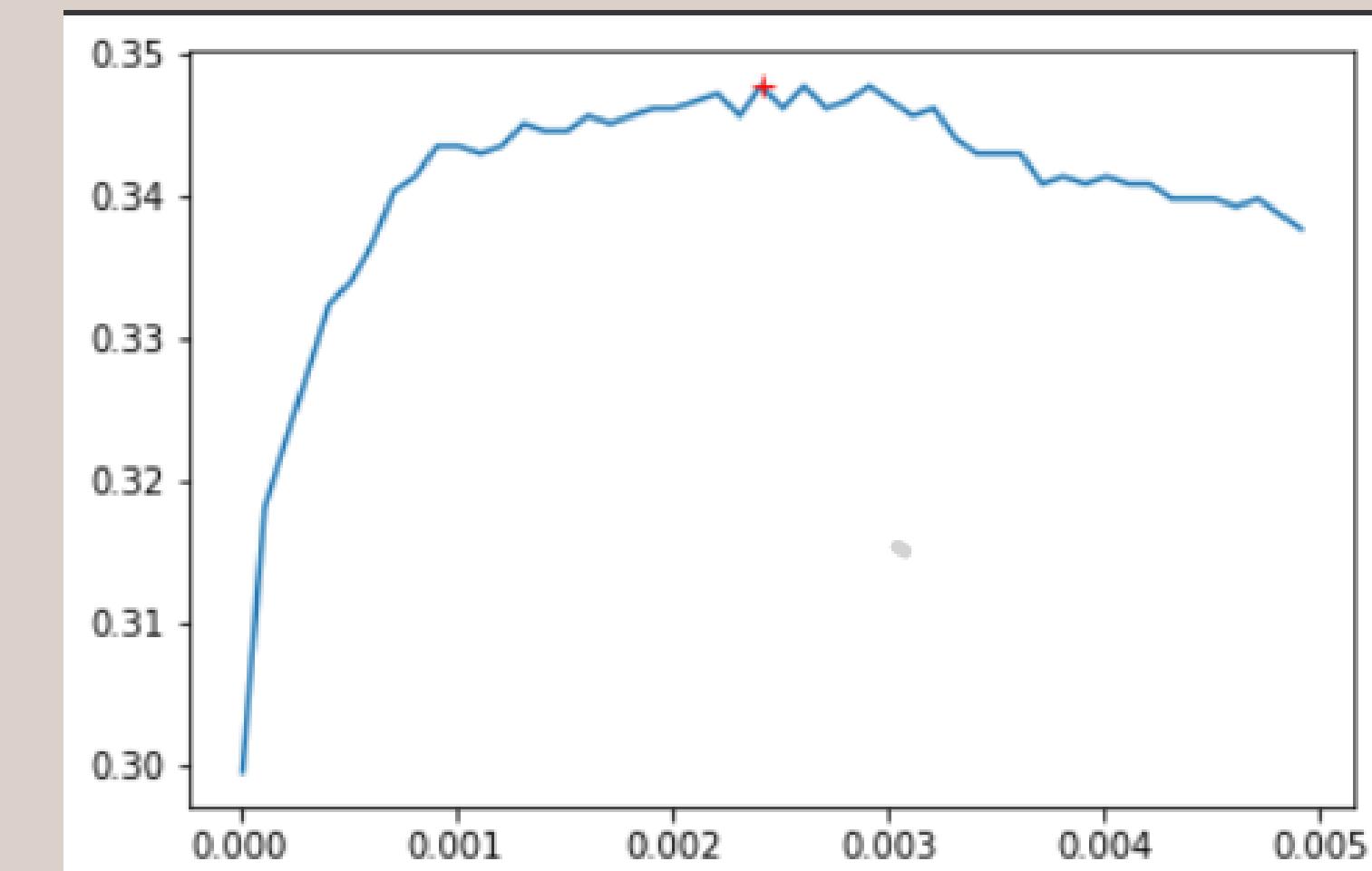


Machine Learning MODELS

The models used for this dataset are:

1. Simple Naive Bayes:

- Train data gives an accuracy of 43.94% and test data gives an accuracy of 34.76%.
- Parameter tuning also does not increase the accuracy significantly.



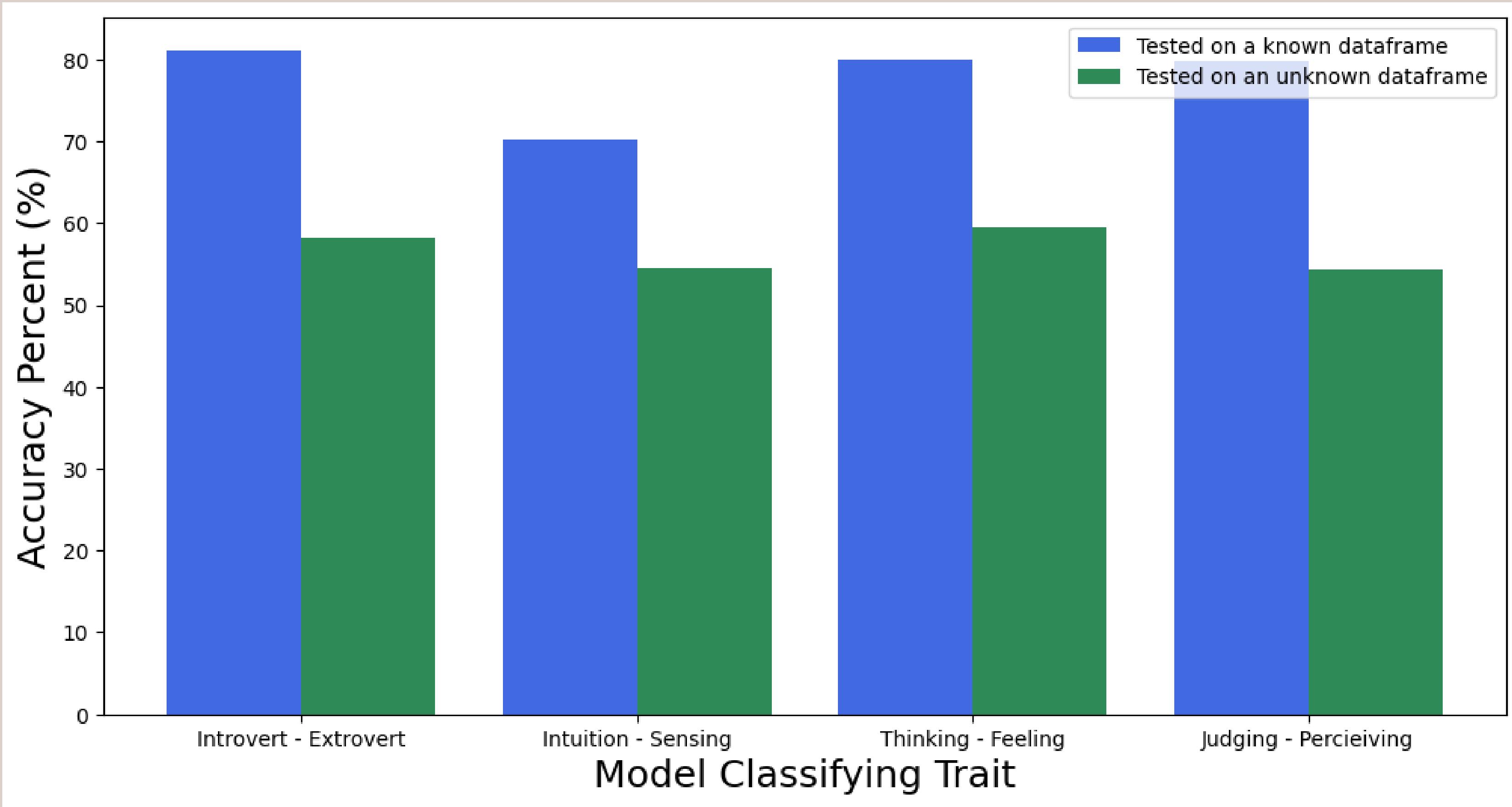
2. SVM Classifier:

- Train Data gives an accuracy of 57.9% which makes it slightly better than Simple Naive Bayes.
- Some of the metrics given by this model are as shown in the figure.

	precision	recall	f1-score	support
INFJ	0.60	0.25	0.35	48
ENTP	0.52	0.38	0.44	143
INTP	0.59	0.38	0.46	53
INTJ	0.56	0.36	0.44	151
ENTJ	0.29	0.22	0.25	9
INFP	0.00	0.00	0.00	8
ENFP	0.00	0.00	0.00	9
ISFP	0.50	0.18	0.27	22
ENFJ	0.56	0.63	0.59	337
ISTP	0.58	0.70	0.64	398
ISFJ	0.58	0.53	0.55	230
ISTJ	0.61	0.65	0.63	285
ESTP	0.58	0.41	0.48	34
ESFP	0.28	0.42	0.33	45
ESTJ	0.38	0.55	0.45	40
ESFJ	0.47	0.61	0.53	75

3.Improvised Naive Bayes:

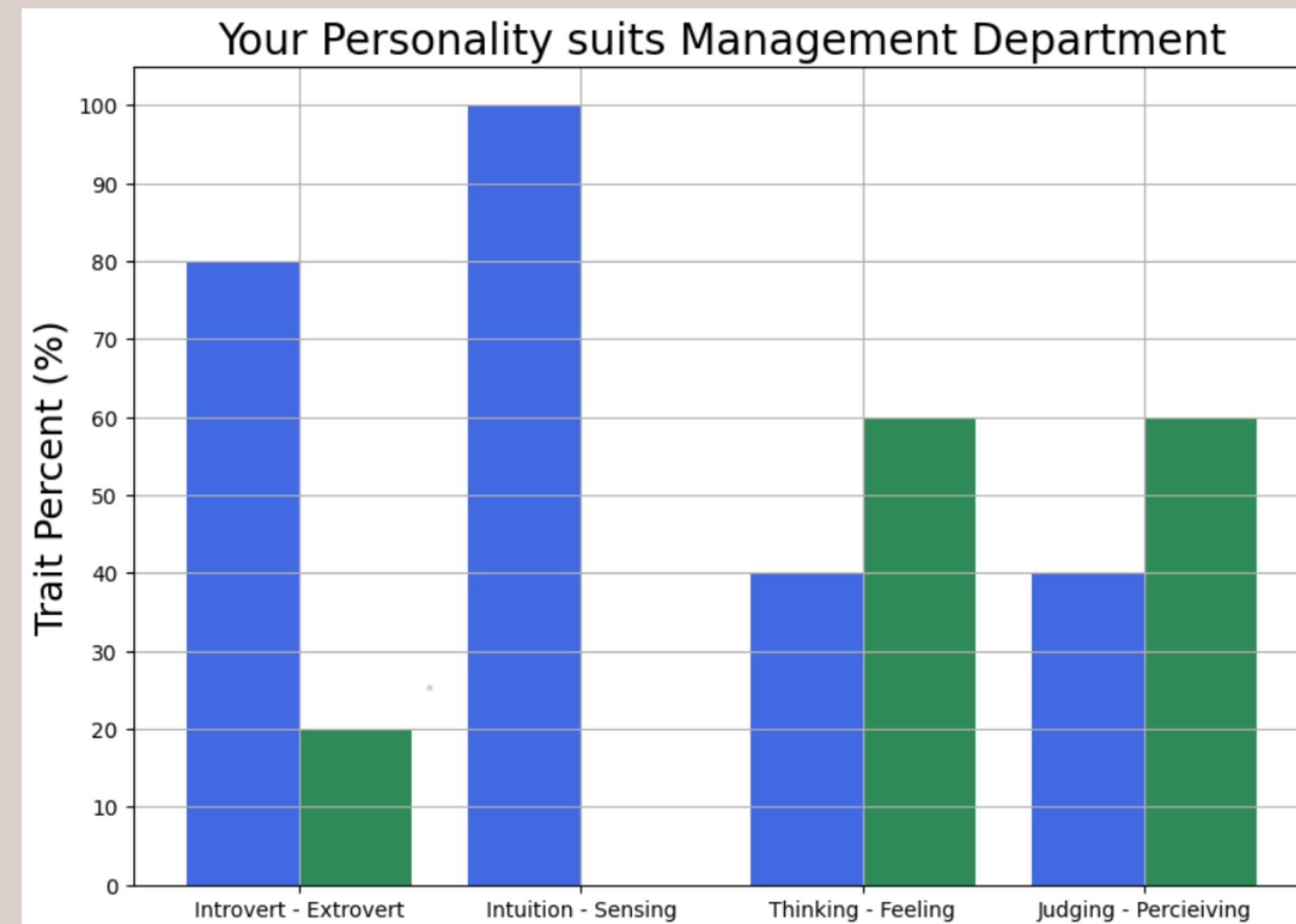
- The Myers Briggs Type Indicator (or MBTI for short) is a personality type system that divides everyone into 16 distinct personality types across 4 axis:
 - Introversion (I) – Extroversion (E)
 - Intuition (N) – Sensing (S)
 - Thinking (T) – Feeling (F)
 - Judging (J) – Perceiving (P)
- In this model, instead of selecting all 16 types of personalities as a unique feature, it will create 4 classifiers to classify each person.
- The accuracy of this model is better than the other two models.



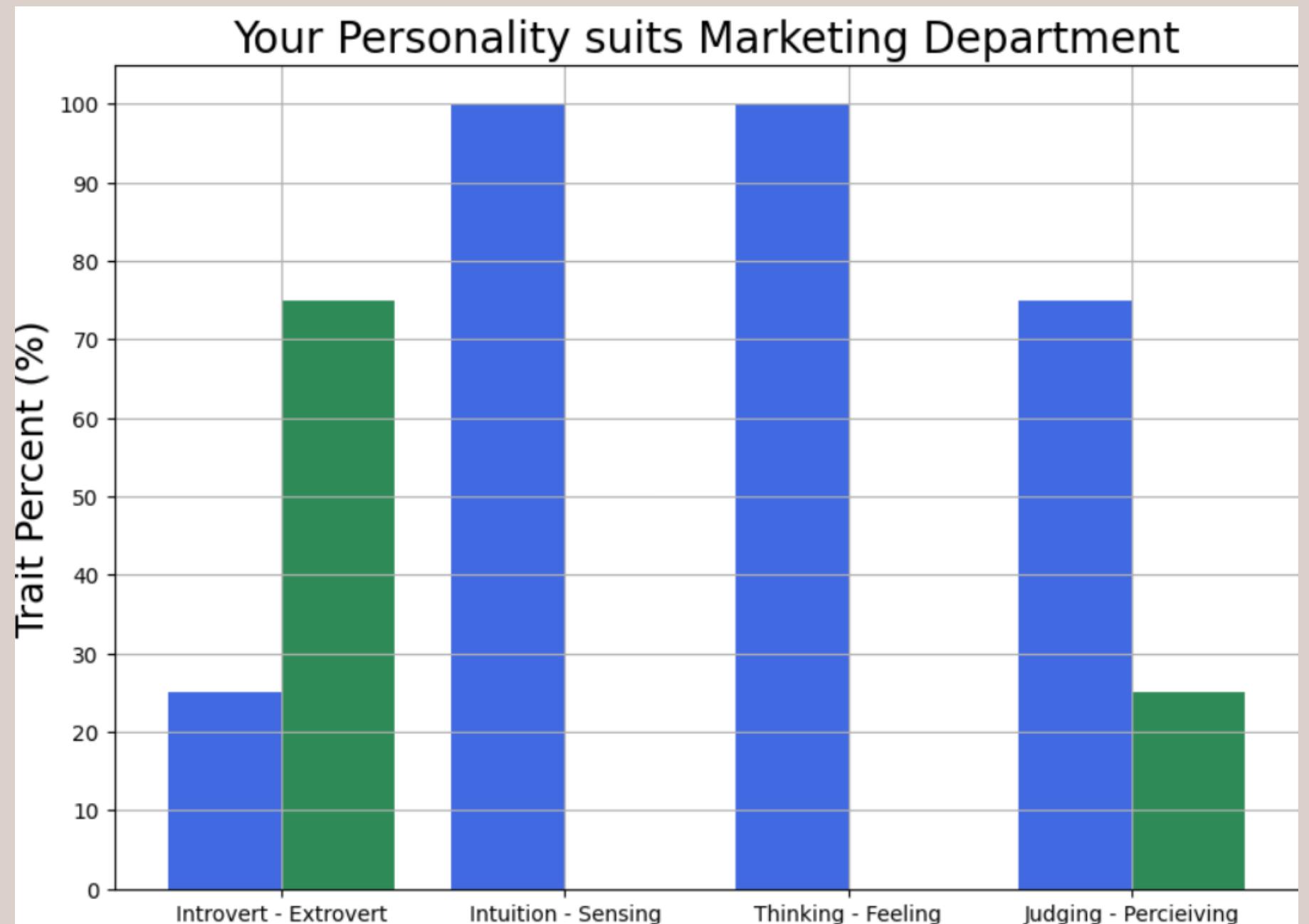
RESULTS

Using Improvised Naive Bayes as our base model, some of the results for club recruitment are:

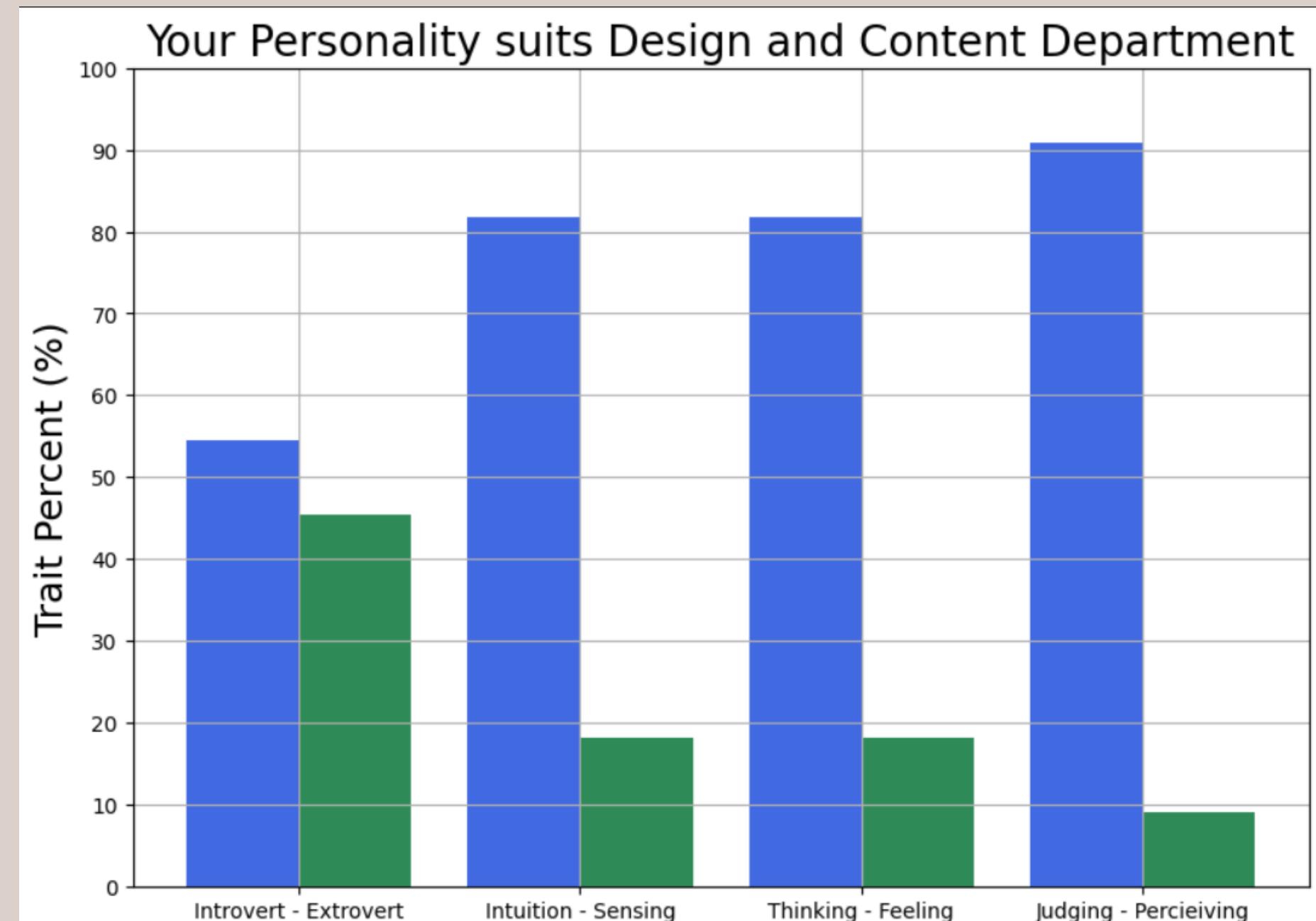
1. Harsha Bhogle



2. Noah Kagan



3. Ken Prewitt



CONCLUSION

- Basic classifiers like Naive Bayes and SVM didn't give expected results.
- Improvised Naive Bayes Classifier provides significant results. Implementing these on a bigger dataset could produce astonishing results which might actually be much better than a human judging someone's personality.
- Currently, this model determines each trait with around 60% accuracy despite the limited dataset.
- This model will be useful for the clubs in VIT to recruit candidates who are fit and to assign the roles that are suitable for them with their social media posts.

REFERENCE

- <https://www.kaggle.com/datasets/datasnaek/mbti-type>

THANK YOU

