Lauren Cardieri
DataMining
Project 2: Personality Test: Extrovert vs Introvert
August 13, 2025

For my classification project, I decided to create a machine learning algorithm that decides if someone is extroverted or introverted depending on numerous factors. These factors include: how many daily hours spent alone, if you feel drained after socializing, if you have stage fright, how many close friends you have, how often you go to social events, how frequently you post on social media, and how frequently you go outside. The findings of this project was interesting, as I found out the features that I thought would make someone extroverted actually made them introverted. My methodology was that this would be an interesting project to conduct for individuals to learn more about themselves.

For example, I discovered that the most important features in determining if someone is extroverted or introverted is whether they have stage fright, how much time they spend alone, and if they were drained after socializing. I was shocked to discover that while friend group size and frequency posting does matter, when coupled with other introverted features, these can actually make you more introverted than extroverted. For example, an introvert may feel more comfortable posting on social media because they are not actually having that interaction face-to-face, thus they do not perceive it as a real interaction. Thus, when coupled with other features such as having stage fright and being drained after socializing, a higher frequency of posting on social media can actually make your score be more introverted! A similar effect occurred with friend group size, to a lesser extent, however it was still interesting to see!

The model I chose for this project was Logistic Regression, as it performed the best on all tests. However, as you can see in Figure 7 and Figure 8, many models performed well on this dataset, with the only one truly struggling being Decision Trees. However, logistic regression had the best accuracy and the second best cross-validation results, so this was the model chosen for this project.

My program can solve a real world problem by helping individuals determine whether they are extroverted or introverted, which will help them discover more about their personalities. There is also a cool feature that allows users to upload a .csv file of their team or classroom's responses to this form. Once uploaded, the model will determine who is introverted and who is extroverted in the class, and what percentages of each they are. Then, if the user uses the team division tab, the model will separate these individuals into teams with equal extroverts and introverts. It will determine a group leader (the most extroverted), a group planner (most introverted), a group communicator (no stage fright and most communication hours), and the rest of the group will be "support". This will hopefully improve productivity in these groups, as a diverse set of personality can help each team grow, as they bring different skillsets and perspectives to the table.

Appendix:

Figure 1:

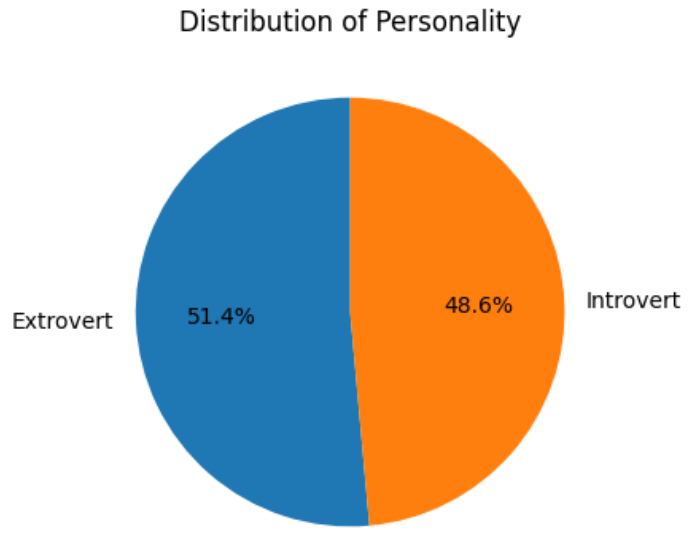


Figure 2:

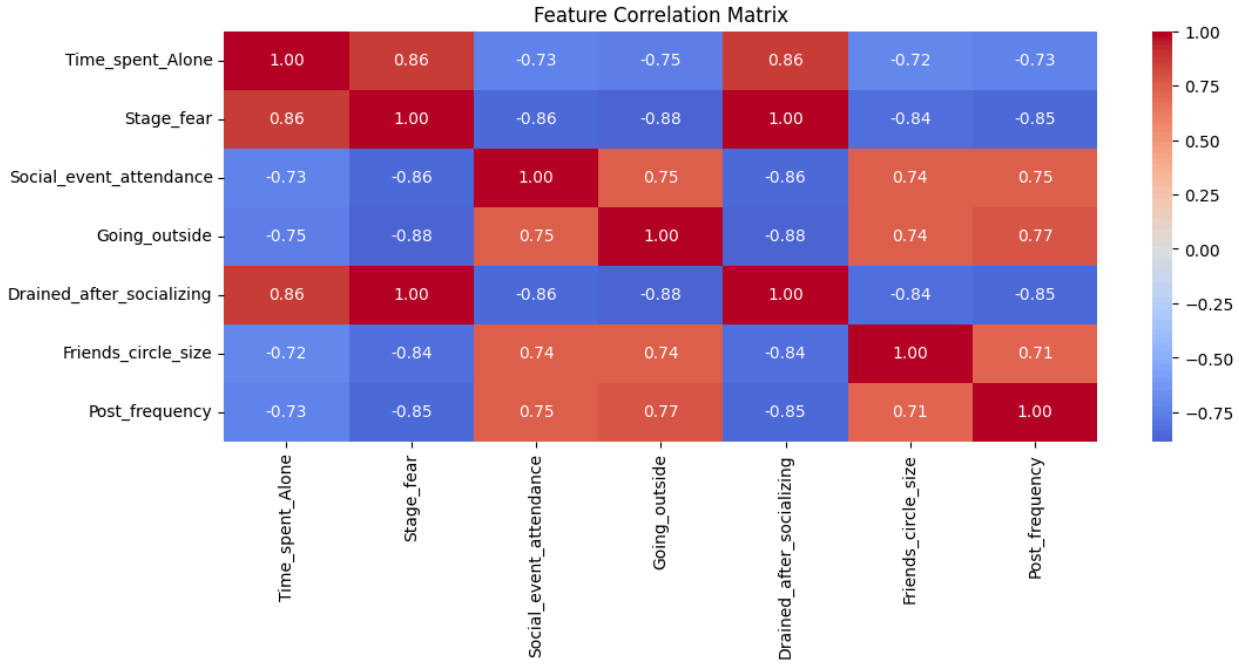


Figure 3:

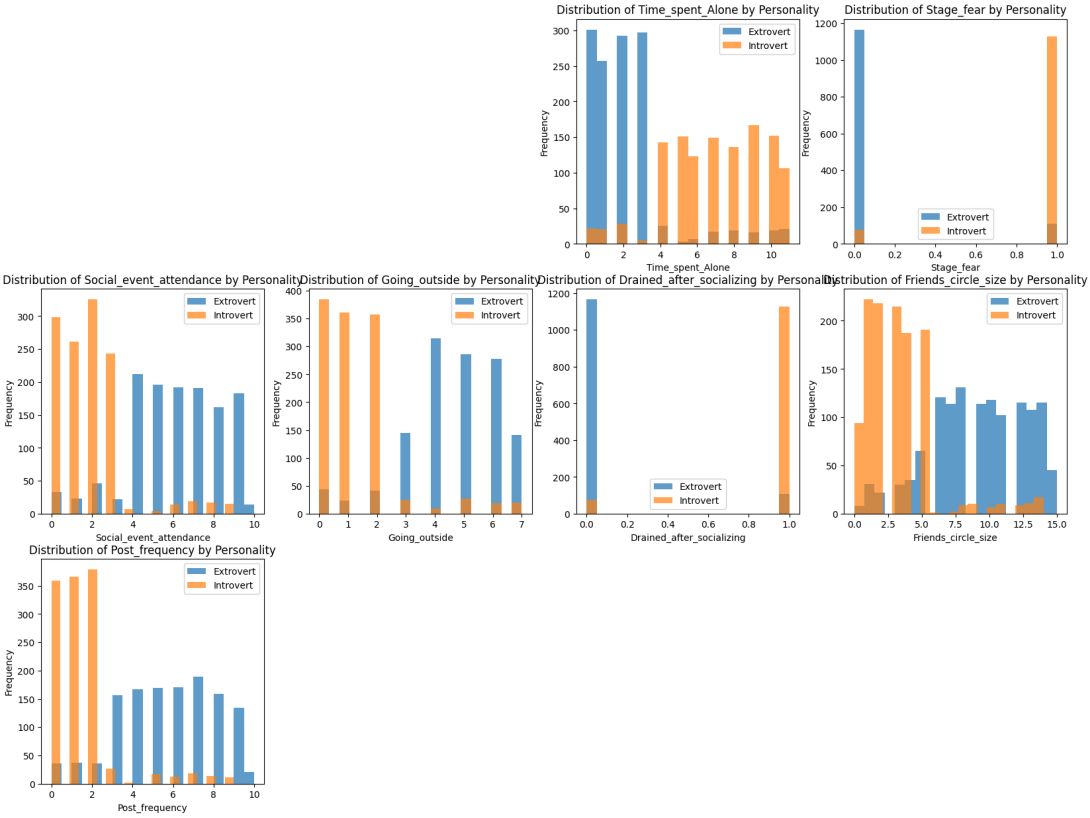


Figure 4:

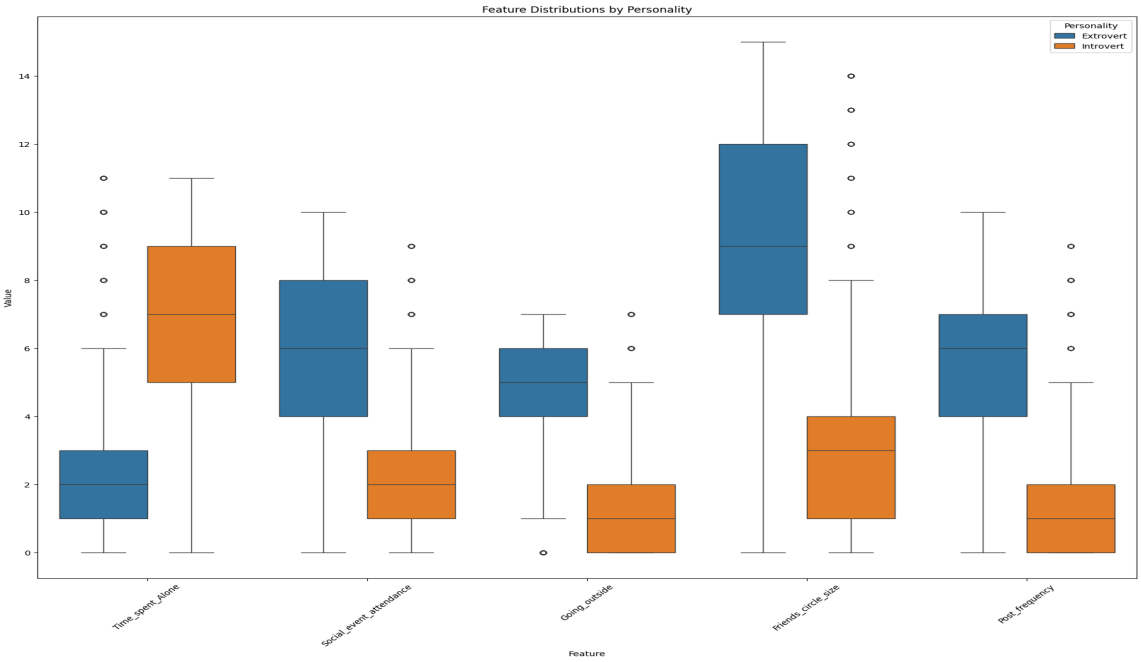


Figure 5:

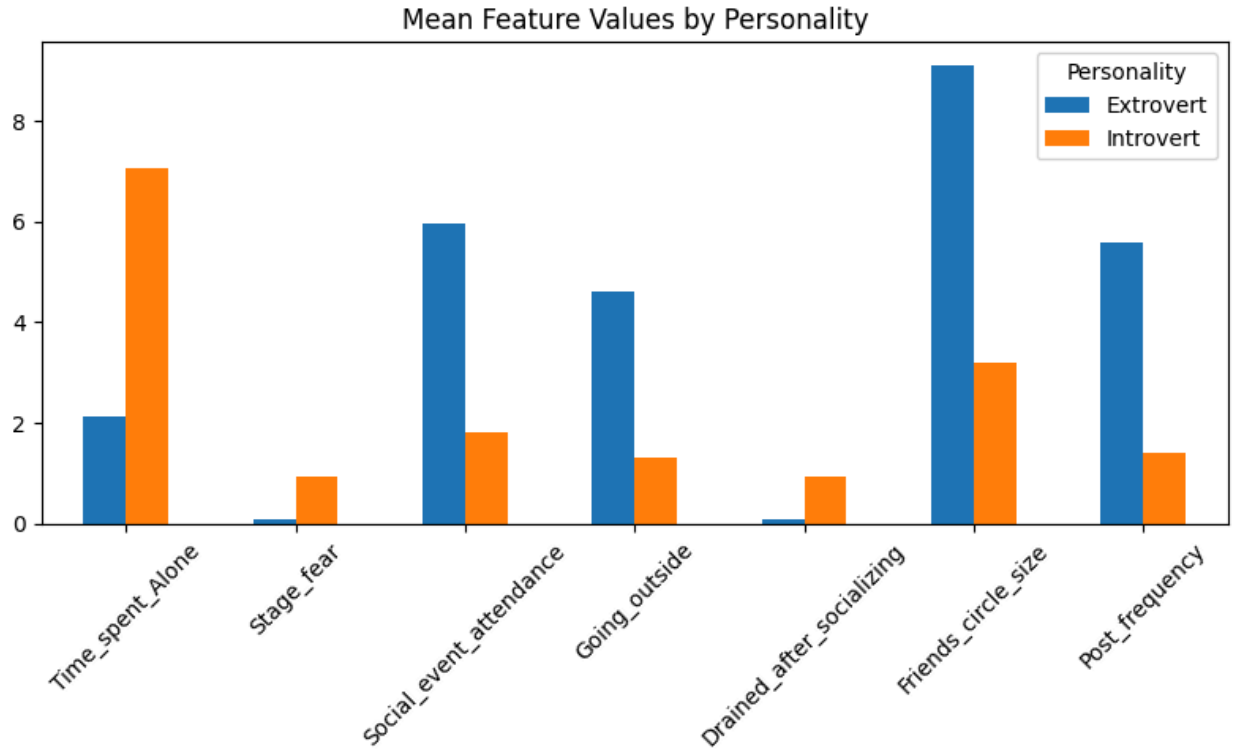


Figure 6:

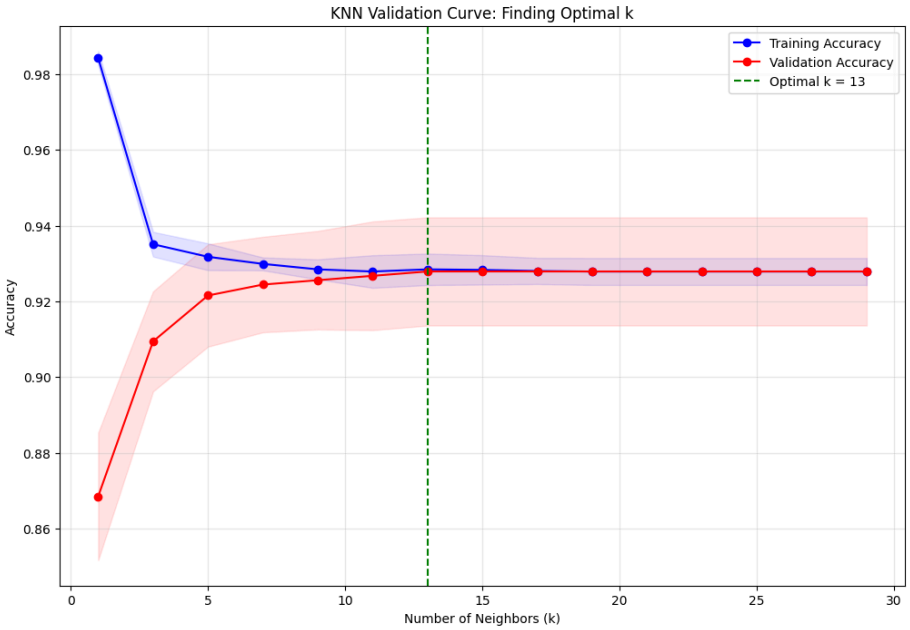


Figure 7:

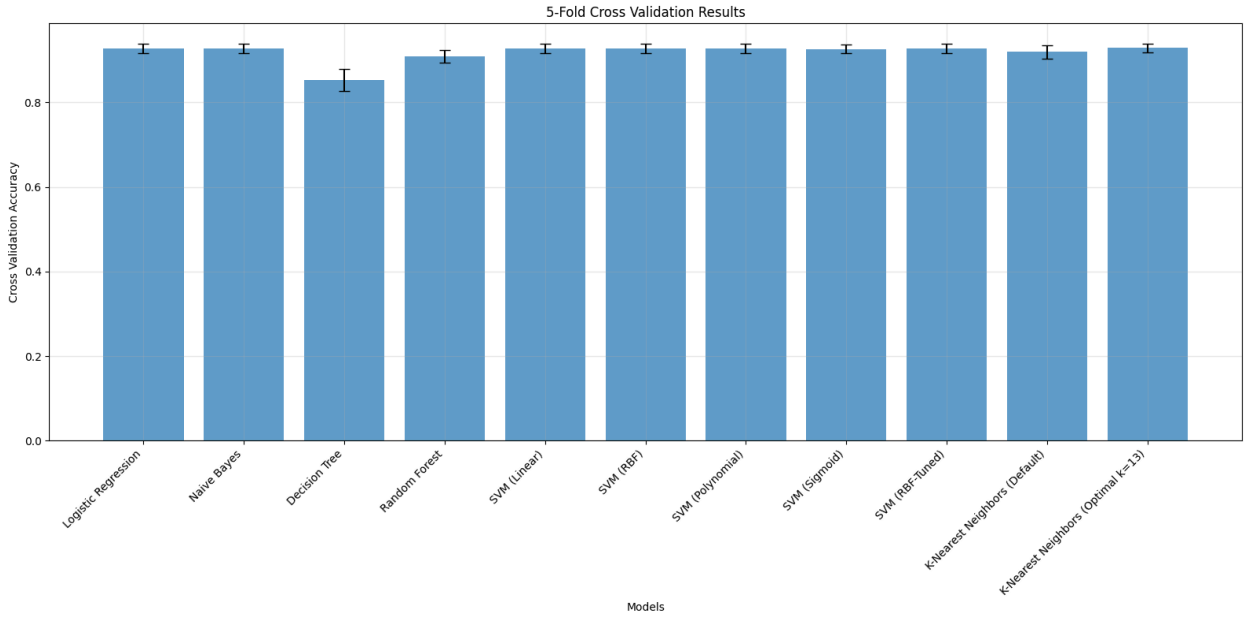


Figure 8:

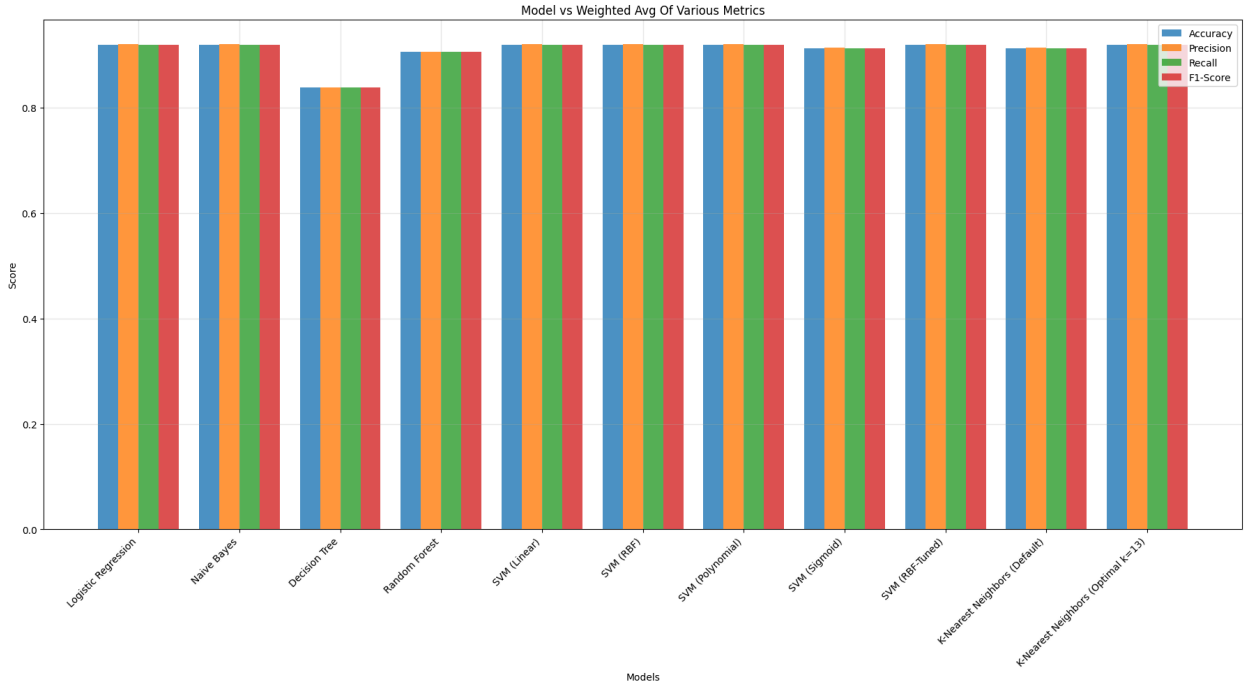


Figure 9:

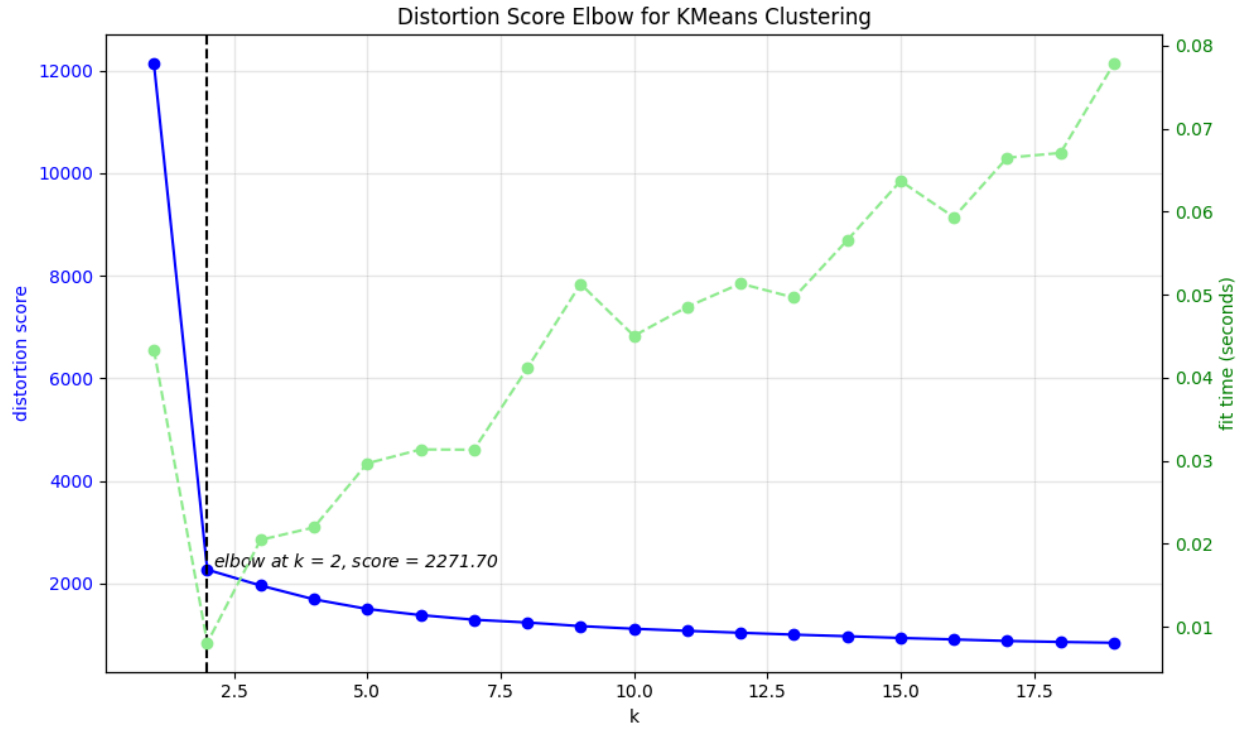


Figure 10:

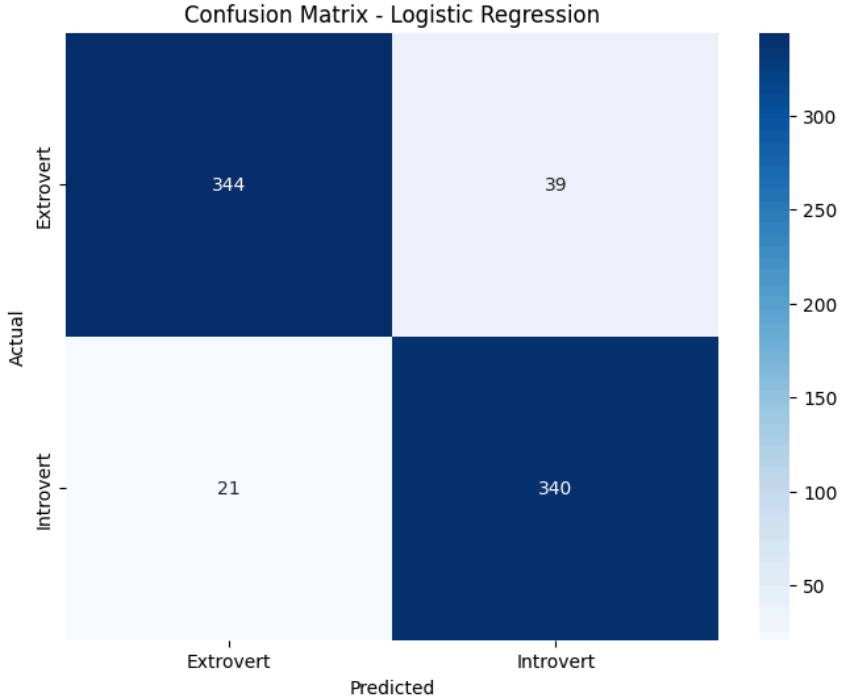
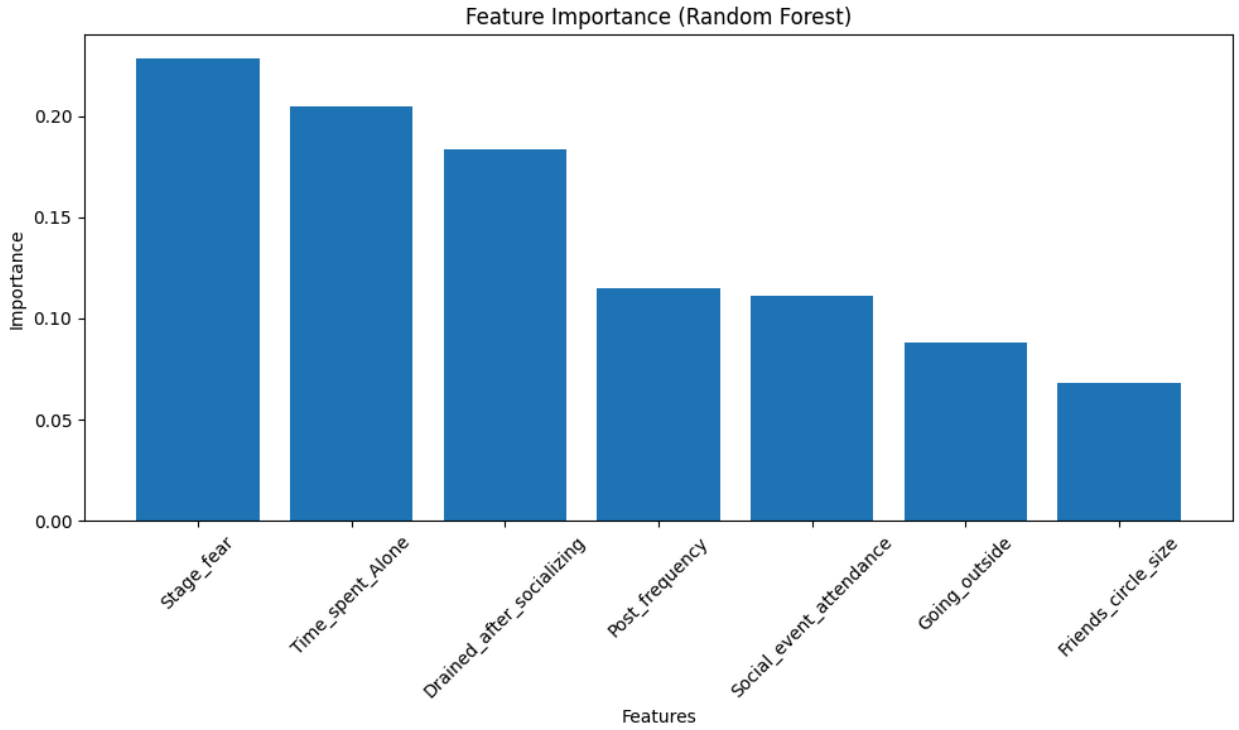


Figure 11:


Feature Importance (Random Forest)
Importance
0.00
0.05
0.10
0.15
0.20
Stage_fear
Time_spent_Alone
Drained_after_socializing
Post_frequency
Social_event_attendance
Going_outside
Friends_circle_size
Features