Results:

TABLE I

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | O | C | E | A | N | Recall |
| Logistic Regression | 0.76 | 0.57 | 0.79 | 0.72 | 0.69 | 0.74 |
| SVM | 0.77 | 0.72 | 0.79 | 0.68 | 0.72 | 0.78 |
| KNN, NB | 0.70 | 0.54 | 0.61 | 0.50 | 0.53 | 0.58 |
| Naïve Bayes | 0.60 | 0.59 | 0.58 | 0.59 | 0.58 | 0.59 |

COMPARISON OF RECALL RATES

TABLE II

COMPARISON OF F1 VALUES

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | O | C | E | A | N | F1 |
| Logistic Regression | 0.77 | 0.63 | 0.79 | 0.71 | 0.70 | 0.75 |
| SVM | 0.77 | 0.72 | 0.79 | 0.67 | 0.70 | 0.72 |
| KNN, NB | 0.61 | 0.54 | 0.56 | 0.50 | 0.52 | 0.54 |
| Naïve Bayes | 0.60 | 0.59 | 0.58 | 0.58 | 0.58 | 0.57 |

TABLE III

COMPARISON OF ACCURACY

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | O | C | E | A | N | Accuracy |
| Logistic Regression | 0.78 | 0.65 | 0.78 | 0.72 | 0.71 | 0.74 |
| SVM | 0.78 | 0.74 | 0.78 | 0.68 | 0.71 | 0.72 |
| KNN, NB | 0.62 | 0.54 | 0.55 | 0.52 | 0.51 | 0.55 |
| Naïve Bayes | 0.61 | 0.62 | 0.57 | 0.55 | 0.562 | 0.59 |

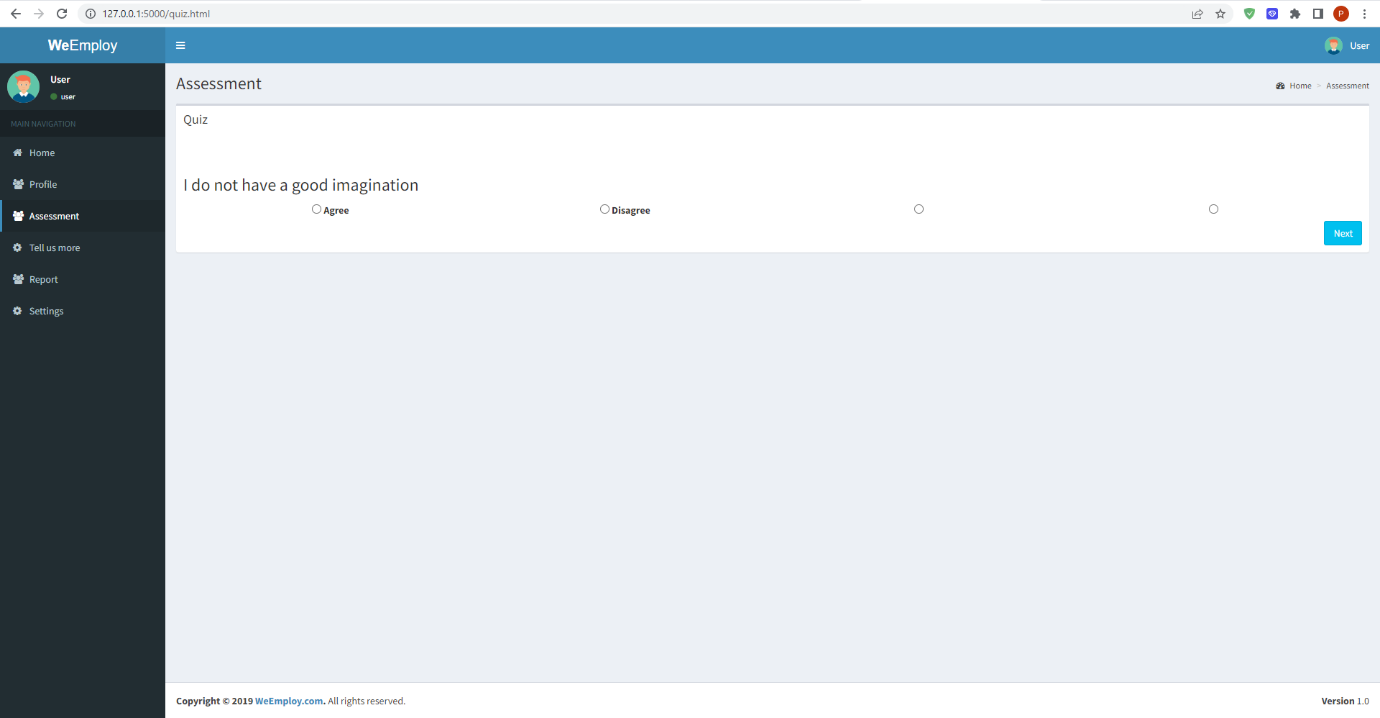
We have Trained our model on different algorithms, Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Naïve Bayes and performed various performance measures such as F1 Score, Recall and Accuracy and compared them with different algorithms. Columns O, C, E, A, N are the predicted values of openness, conscientiousness, extraversion, agreeableness and neuroticism. In Table 1 we have recall, SVM has the highest recall score of the 4 algorithms of 78% . In Table II Logistic Regression has the highest F1 VALUES of the 4 algorithms of 75%. In Tables III Logistic Regression has the highest accuracy of the 4 algorithms of 72%. If we average all the scores of 3 performance measures of 4 algorithms, we find that Logistic regression has the highest average so we have used linear regression as our model predicting the personality. Second highest average is SVM it performs good in situations where the data is imbalanced and one class data is less than other class.

There are several limitations to using machine learning for personality prediction in the context of CV analysis. One of the main limitations is that personality is a complex and multifaceted construct, and it can be difficult to accurately predict it based on limited information such as that provided in a CV or cover letter. Machine learning algorithms may not be able to capture the full range and depth of personality traits and may be prone to making errors or oversimplifying the data.

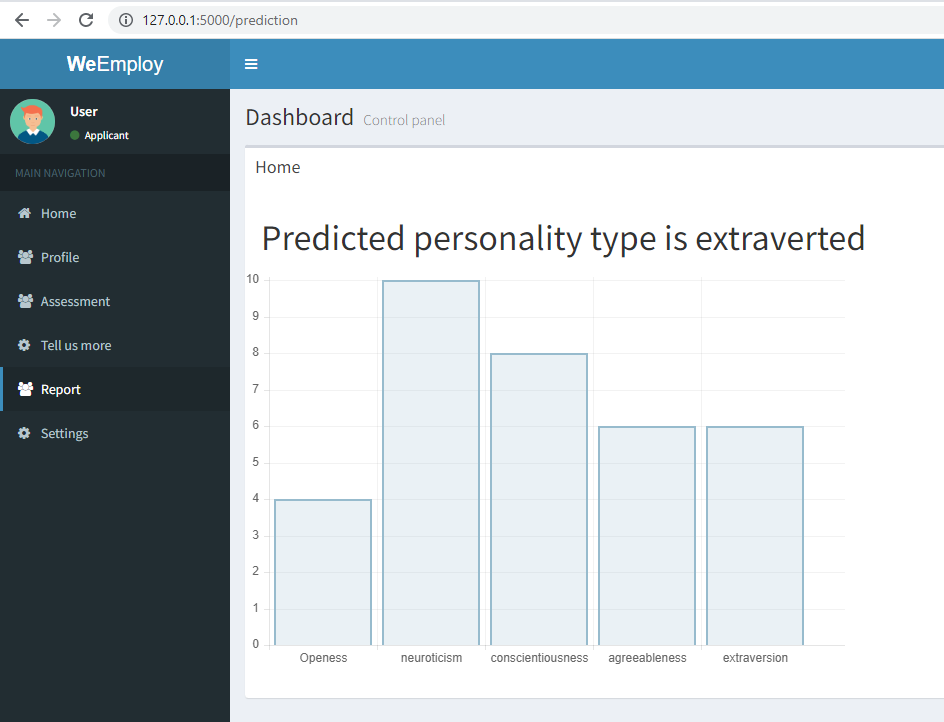
Another limitation is that machine learning algorithms can be biased if the training data contains biases. There is also a risk that personality prediction could be used to discriminate against certain groups of people.

It is important to keep in mind that personality is just one factor that should be considered in a CV analysis. Other factors, such as skills, experience, and education, are also important and should be given due consideration.

Questionnaire:



Results:



Results are shown after answering the questionnaire.