APPROCH FOR PRIDICTING EMPLOYEE ATTRITION

Given a initial training data set of 19104 rows, 13 columns & test data set of 741 rows,1 columns it was eminent that test data needs to be mapped with test data set for algorithm to work.

Also, for Attrition value a conversion function was needed to convert it to '1' -> left the job & '0' -> working & term duration is required to be calculated to get a understanding of employees long standing relationship with the company.

Name of the column was not as per required syntax & renaming needs to be carried out.
Some outliers got detected for term duration which was showing negative which has been handled.
After dataset cleaning it is required to segregate numerical columns from categorical column & perform 'one hot encoding' on categorical columns & then concatenating numerical column with categorical one hot encoded column.

This makes our data set ready to be modelled.
Now, it is time to separate features from target & our target being Attrition is separated out & 20% data is set aside for validation.

Five different algorithms namely 'Linear Regression', 'Ridge Regression', 'Lasso Regression', 'K Neighbors Regressor', 'Decision Tree Regressor' is tested & compared using Root Mean Square Error as to see which is the most compatible algorithm.

Now, test data needs to be mapped with training as mentioned earlier to apply final prediction of test data.

Lastly, f1_score is calculated using precision & recall to summarize the accuracy of the code.

With f1-score :1
Accuracy:1
Macro avg:1
This code looks perfect to predict employee attrition.