Steps to be performed are as follows:

1) Import required libraries - 1 Mark

2) Read dataset and perform Text processing for the tweets ( Remove Stop words , special characters and convert the text to lowercase etc. ) - 2 Mark

3) Using the train\_test\_split function of Sklearn, Split train and test dataset - 1 Mark

4) Create pipeline and define parameters for GridSearch ( You might Refer the code below ) - 1 Mark

text\_clf = Pipeline([('vect', CountVectorizer()),  
                     ('tfidf', TfidfTransformer()),  
                     ('clf', MultinomialNB())])

tuned\_parameters = {  
    'vect\_\_ngram\_range': [(1, 1), (1, 2), (2, 2)],  
    'tfidf\_\_use\_idf': (True, False),  
    'tfidf\_\_norm': ('l1', 'l2'),  
    'clf\_\_alpha': [1, 1e-1, 1e-2]  
}

5) Perform classification (using GridSearch) - 2 Mark

6) Print the confusion matrix, accuracy, F1 score on the **test dataset** - 2 Mark

7) Comment about the classification metric values you have obtained in part 6 - 1 mark

8) Try a different set of hyperparameters (defined in tuned\_parameters) and comment on its effect on resultant classification metrics. - 2 mark