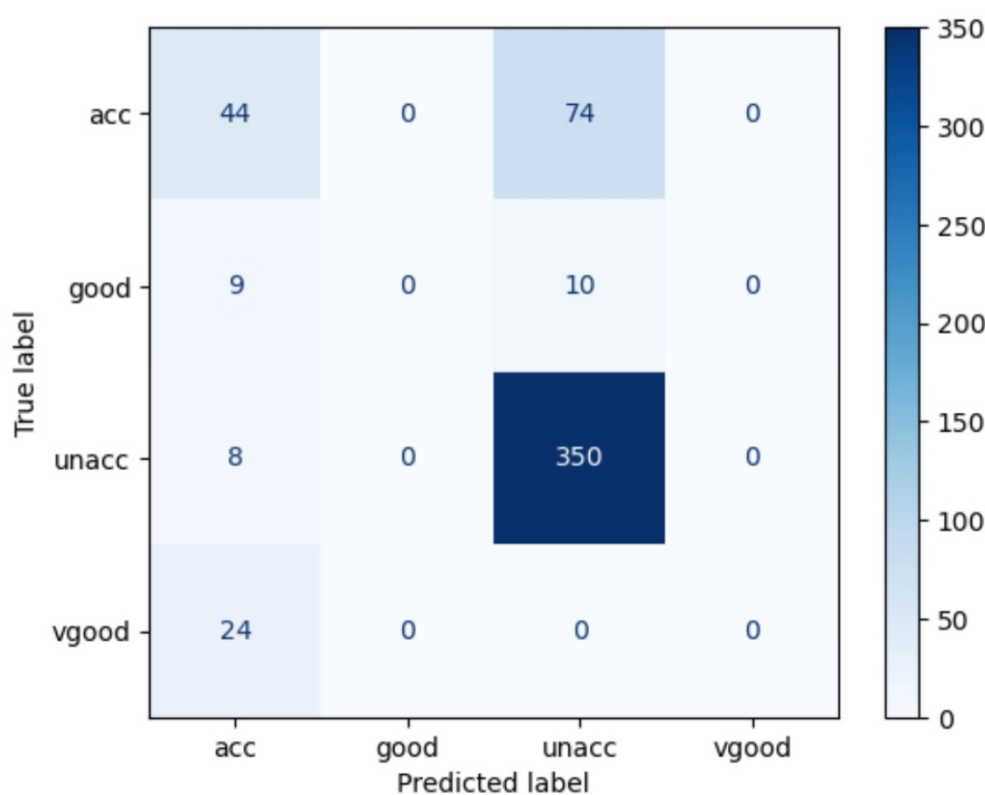Q1: Provide the confusion matrix and explain the classification model performance. According to 'sklearn documentation' Wikipedia and other references may use a different convention for axes. To make sure your interpretation is correct use the result of Task 2.3 as a guide.

```
Confusion Matrix:
[[ 44    0   74    0]
 [  9    0   10    0]
 [  8    0  350    0]
 [ 24    0    0    0]]
```



The "acc"class has total of 118 instances, 44 were successfully predicted and belong to "acc"class, 74 instances that were mistakenly predicted and belongs to "unacc"class. For the "unacc"class, there were 350 instances that were correctly predicted as belonging to the "unacc" class, only 8 instances were erroneously classified as "acc" class. For "vgood" and "good" class, all the instances were incorrectly classified.

Q2: Provide the evaluation metrics including 'accuracy, precision, recall, and f1-score). Explain your findings.

|              | precision | recall | f1-score | support |
|-------------:|----------:|-------:|---------:|--------:|
| acc          | 0.52      | 0.37   | 0.43     | 118     |
| good         | 0.00      | 0.00   | 0.00     | 19      |
| unacc        | 0.81      | 0.98   | 0.88     | 358     |
| vgood        | 0.00      | 0.00   | 0.00     | 24      |
| accuracy     |           |        | 0.76     | 519     |
| macro avg    | 0.33      | 0.34   | 0.33     | 519     |
| weighted avg | 0.67      | 0.76   | 0.71     | 519     |

"Unacc"class has the highest precision, which is 0.81, also "unacc"class has high recall, which means the model correctly identifies a large proportion of positive instances and avoids misclassifying them as negative.

The F1-score is a metric that combines precision and recall into a single value, providing a balanced measure of a model's performance. "Unacc"class got highest f1-score 0.88， which indicating better performance of the model in terms of both precision and recall. The accuracy for overall measure of correct prediction is 0.76.