Task1: Chained multi-outputs.

In the given dataset, we have three classes: Type2, Type3, and Type4. The accuracy of a prediction model depends on the correct prediction of Type2 first, and then on the subsequent correct predictions of Type2 + Type3 and finally Type4.

For example, we developed a model (M), and we are evaluating the model with three instances in the test samples. For simplicity, we assume that Type2, Type3, and Type4 can have only two values, Yes and No.

Email Instance 1:

- True Labels: Type2: Yes, Type3: Yes, Type4: Yes
- Predicted Labels: Type2: Yes, Type3: Yes, Type4: Yes
- Evaluation:
    - Type2: Correct
    - Type2 + Type3: Correct
    - Type2 + Type3 + Type4: Correct
    - Final Accuracy for Instance 1: 100%

Instance 2:

- True Labels: Type2: Yes, Type3: No, Type4: Yes
- Predicted Labels: Type2: Yes, Type3: Yes, Type4: No
- Evaluation:
    - Type2: Correct
    - Type2 + Type3: **Incorrect** (since Type3 is incorrectly predicted)
    - Type2 + Type3 + Type4: Not evaluated since ***Type3 is wrong***
    - Final Accuracy for Instance 2: 33%

Instance 3:

- True Labels: Type2: No, Type3: Yes, Type4: Yes
- Predicted Labels: Type2: Yes, Type3: Yes, Type4: Yes
- Evaluation:
    - Type2: Incorrect
    - Type2 + Type3: ***Not evaluated since Type2 is wrong***
    - Type2 + Type3 + Type4: ***Not evaluated since Type2 is wrong***
    - Final Accuracy for Instance 3: 0%

Instance 4:

- True Labels: Type2: No, Type3: Yes, Type4: Yes
- Predicted Labels: Type2: No, Type3: Yes, Type4: No
- Evaluation:
    - Type2: Correct
    - Type2 + Type3: *Correct*
    - Type2 + Type3 + Type4: ***Incorrect***
    - Final Accuracy for Instance 4: 67%

The key point in this evaluation strategy is that the accuracy of predicting Type3 and Type4 doesn't contribute to the final accuracy unless Type2, Type2+Type3 is correctly predicted. To calculate the final accuracy of the system, you'd average the accuracy of each instance. In the case of our three instances:

Total Accuracy = (100% + 33% + 0% + 67%) / 3 ≈ 50.33%

Task2: Hierarchical Modelling

It is a multi-stage approach to classification, where each stage filters the data for the subsequent model based on the predictions of the previous model. In the given dataset, we have three class variables, Type2, Type3, and Type4. In the first stage, all the data would be processed through a model (M1 would be created) to classify the emails based on Type2. In the Type2, there are multiple values, i.e., others, Problem/Fault, and Suggestion. Thus, M1 would classify instances in those classes.

In the next stage, a filtered instances based on prediction of M1 should be selected for the development of new model, for instance M12. As you can see, in the below diagram, newly created models are based on the classification values of M1.

<table>
<tr><td>Dataset</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="12"></td><td rowspan="12">M1 (Type2)</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Others</td><td>M12(Type3)</td><td></td><td></td></tr>
<tr><td rowspan="3"></td><td rowspan="3"></td><td>C1</td><td>M121(Type4)</td></tr>
<tr><td></td><td rowspan="2"></td></tr>
<tr><td>C2</td><td>M122(Type4)</td></tr>
<tr><td>Problem/Fault</td><td>M13(Type3)</td><td></td><td></td></tr>
<tr><td rowspan="3"></td><td rowspan="3"></td><td>C1</td><td>M131(Type4)</td></tr>
<tr><td></td><td></td></tr>
<tr><td>C2</td><td>M132(Type4)</td></tr>
<tr><td>Suggestion</td><td>M14(Type3)</td><td></td><td></td></tr>
<tr><td></td><td></td><td>C1</td><td>M141(Type4)</td></tr>
</table>

				C2	**M142(Type4)**

In this case, one way of computing accuracy of the system would be to find out:

- **Correct Classification at All Stages (Perfect Path):** The instance is correct at **Type2**, **Type3**, and **Type4**. This is a full score for that instance.
- **Partial Correctness:** If an instance is correct at **Type2** but incorrect at **Type3**, the chain stops, and the instance receives partial credit based on the number of correct stages.
- **Incorrect at First Stage (Type2):** If an instance is incorrect at **Type2**, it receives a score of zero, as subsequent classifications are based on the first being correct.