### Model Results Summary and Insights

#### 1. Actual vs Predicted Class Distribution

**Insight:** - “Fresh” is most frequently predicted, even when the actual class is different. - “Late\_Spoilage” is often predicted for samples that are actually “Spoiling”.

**Implication:** - Indicates misclassification between closely related stages. - Suggests imbalance and overlapping feature space in intermediate spoilage stages.

#### 🫮2. Confusion Matrix Summary

| True  Predicted | Fresh | Spoiling | Late\_Spoilage |
| --- | --- | --- | --- |
| **Fresh** | 629 | 0 | 126 |
| **Spoiling** | 15 | 74 | 317 |
| **Late\_Spoilage** | 0 | 0 | 15 |

**Insight:** - High accuracy on “Fresh” - “Spoiling” is frequently confused with “Late\_Spoilage” - True positive rate for “Late\_Spoilage” is good but with very low support

**Implication:** - Suggests need for feature enhancements to better separate middle spoilage stage

#### 3. Multiclass ROC Curve

| Class | AUC Score |
| --- | --- |
| **Fresh** | 0.96 |
| **Late\_Spoilage** | 0.95 |
| **Spoiling** | 0.57 |

**Insight:** - Model is confident and accurate for extreme classes - Weak in identifying the intermediate class “Spoiling”

**Implication:** - Transition zone around spoilage onset is biologically complex - Possible need for more granular or time-aware features

#### 4. Correct vs Incorrect Predictions by Class

**Insight:** - “Fresh”: high correct prediction rate - “Spoiling”: majority incorrectly predicted as “Late\_Spoilage”

**Implication:** - Transitional phase needs further discrimination through engineered features

#### 5. XGBoost Feature Importance

| Feature | Relative Importance |
| --- | --- |
| mean\_BME\_Temp | High |
| mean\_BME\_VOC\_Ohm | High |
| mean\_MQ3\_Top\_PPM | High |
| VOC\_to\_MQ3\_Bottom\_Ratio | Moderate |

**Insight:** - Temperature, VOC, and ethanol levels are core indicators of spoilage stages - Delta features are weak likely due to short window size

**Implication:** - Incorporate longer temporal windows or trend-based features - Potential benefit from LSTM or time-series-aware model

*6. ROC Curves (Multiclass)*

**Image**: multiclass\_roc\_curve.png

| **Class** | **AUC Score** |
| --- | --- |
| **Fresh** | 0.96 |
| **Late\_Spoilage** | 0.95 |
| **Spoiling** | 0.57 |

**Insight**:

* High AUC for "Fresh" and "Late\_Spoilage": model is highly confident in predicting extreme ends.
* Poor AUC for "Spoiling": model is weak at distinguishing this middle class.

**Implication**:

* Indicates that spoilage progression is more easily classified at beginning and end phases.
* "Spoiling" stage might require **more distinctive biomarkers** or time-based features for separation.

### Final Notes

* The model captures biological relevance of spoilage — e.g., fermentation-induced ethanol and VOC rise.
* Overall accuracy ~60% with high precision for extremes but low performance on mid-class.
* Model performance can be enhanced with:
  + Data augmentation or balancing (for “Spoiling”)
  + Feature engineering with temporal dynamics
  + Possibly expanding training set diversity