

Problem 2: Quality control in food production

A large food manufacturing company, concerned about maintaining high standards in its products, is implementing a new quality control measure to ensure the safety and integrity of its food items. Instead of relying solely on traditional laboratory testing methods, which can be time-consuming and costly, the company aims to introduce a preliminary sensory-based test using taste, smell and visual indicators. This sensory test will serve as an initial screening to identify products with abnormal taste or odor characteristics that may indicate quality issues. Subsequently, those flagged by the sensory test will undergo more comprehensive and accurate laboratory testing.

The outcome of the test reports the value of the variables: **taste**, which quantifies the perceived intensity of flavors in the food; **odor**, representing the strength of any odors emitted by the food; **smooth**, indicating the smoothness or roughness of the food's texture; and **bright**, measuring the brightness or vividness of the food's color. Additionally, the variable **result** indicates whether these products passed or failed the laboratory quality tests.

The aim is to build a classifier for discriminating between high-quality and low-quality food products based on the four sensory parameters listed above.

The cost associated with the laboratory quality tests is \$500 per product. This cost is equivalent to the estimated economic benefit of correctly identifying a true low-quality product. The economic loss incurred due to failing to detect a low-quality product is quantified as a cost of \$100,000. The company estimates that 0.1% of food products have quality issues.

- a) Which classification method should be used? Verify the underlying assumptions.
- b) Build the corresponding classifier, providing its apparent error rate and an estimate of its actual error rate through leave-one-out cross-validation. What do you observe? Which characteristic of the classifier you are using is being highlighted here?

The company intends to use the classifier developed in (b) to identify potential low-quality products during the upcoming production batch, which will consist of 1000 food items.

- c) How much should be budgeted for the cost of the laboratory quality tests?
- d) Compared to the previous strategy of conducting accurate laboratory tests on all products, does the new two-fold testing approach lead to a lower expected cost of misclassification? How much would be saved for this upcoming production batch?

Upload your results here: <https://forms.office.com/e/4btXgcZA8x>