Milk Grading system using machine learning

A milk grading system is important for ensuring the quality and safety of milk products. It helps to assess and classify milk based on various factors such as composition, purity, and hygiene standards. The grading system provides valuable information to consumers, regulators, and manufacturers, enabling them to make informed decisions about milk quality, processing, and product selection. It plays a crucial role in maintaining industry standards, promoting consumer confidence, and ensuring the availability of safe and high-quality milk products in the market.

Feature engineering

This dataset does not contain any missing values.It contains 4 categorical features namely Taste,odor,turbidity and fat. The dependent variable which is grade is a also a categorical feature.Making this a classification problem.Ph and Temperature are continous variables.pH has a lot of outliers. 169 values are below the 25th percetile and 210 values are above the 75th percetile.Further univariate analysis of the features revealed that the Colour coloumn had imbalance with nearly 600 records just being of the value '255'.Grade has 3 dimensions 0.0 ,0.5 and 1.0.But since classification was to be performed, all the 0.5 values were converted to 1.0 which represents that the quality of the milk is satisfactory. Visualization using a heatmap was done to perform bivariate analysis. Odor and Turbidity was highly correlated with each other.

Feature selection

The feature turbidity is removed because odor was correlated with it. This is to perform dimensionality reduction

Model Selection

- Support vector machines: gave an accuracy of 91 percent with the test size being 20% and random state being 42
- Logistic regression : gave an accuracy of 86 percent with the precision of negative class being 0.85 and positive class being 0.86
- decision trees: Gave an accuracy of 99 percent with the test size being 20%

Classification report of decision trees

An accuracy of **99** percent was been acheived using decision trees classifier

```
In [32]: print(classification_report(y_test, y_pred))

              precision   recall  f1-score   support

           0      0.99      0.99      0.99        78
           1      0.99      0.99      0.99       134

    accuracy                          0.99       212
   macro avg      0.99      0.99      0.99       212
weighted avg      0.99      0.99      0.99       212
```

Running the application

1. From the root directory of the project type " python model.py "
2. This would create a pickle file named " model.pkl"
3. Run " python app.py"
4. This would start the application on localhost at port 3000