PROJECT REPORT

Ingredient Analyser for Consumables

Prepared by: Swathi
(swa094@gmail.com)

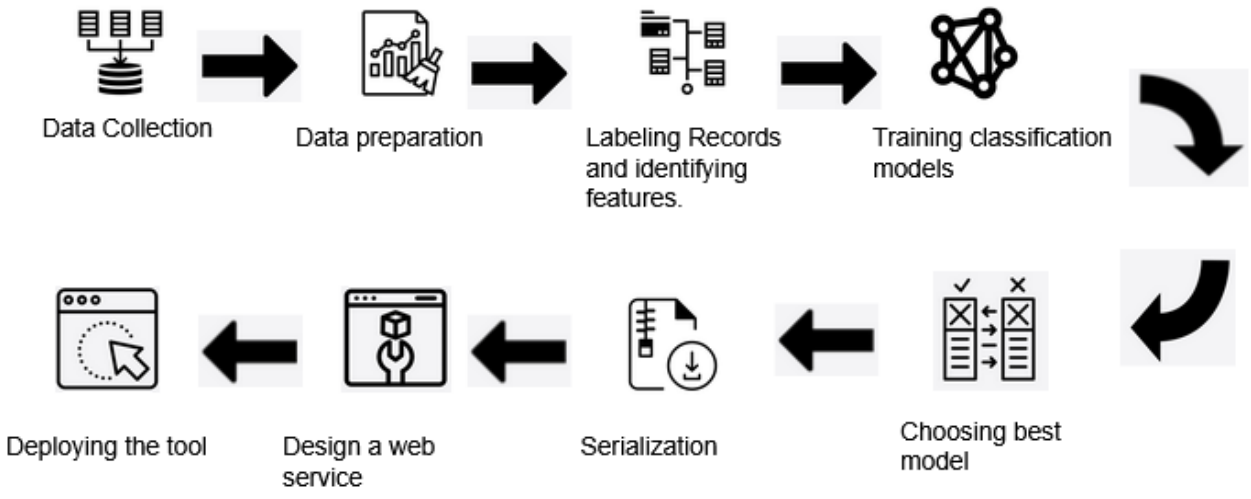
Abstract

Food Additives that are present in certain packed consumable products cause unwanted side effects if they are present in a large quantity or if consumed on a regular basis, especially to children and people with underlying disorders like thyroid or heart conditions. This project aims to develop a tool which uses Machine Learning principles to classify consumables based on their ingredient lists thereby helping customers choose safer products.

The tool uses a classification model to classify the product into one of 5 classes on a scale of 1 to 5 based on the relative harmfulness of the additives, 5 being the safest class.

Methodology:

The Project involved the below stages:



1. **Data Collection**
   A collection of 10000 consumable products with their full ingredient list was obtained.

2. **Data Preparation**
   Unwanted features were dropped and missing values were handled.

3. **Labeling Records**
   Each Record is labeled intuitively on a scale of 1 to 5 based on the position and importance of the ingredients. Features are extracted from the ingredient list to train the ml model.

4. **Training Classification Models**
   The data is split into training and testing subsets. Logistic Regression, Naive-Bayes Classifier, K-Nearest Neighbors, SVM, Random Forest Classifier and Decision Tree Classifier models are used to build classification models on the training data set.

5. **Choosing the best Model**
   Performance metrics of the classification models are analyzed to choose the best model.

6. **Serialization/Pickling**
   The trained model is saved into a file using Pickle module.

7. **Designing a Web Service**
   A simple flask application is developed to take inputs from the user and provide the classification.

8. **Deploying**
   An HTML interface is built and hosted for ease of use.

Results:

K-Nearest Neighbors, SVM, Random Forest Classifier and Decision Tree Classifier models had a significantly higher accuracy rate compared to Logistic Regression and Naive Bayes Classifier models.

The web service created is hosted on pythonanywhere and can be accessed at Swathi071094.pythonanywhere.com.

Steps to use the tool:

1. Copy the ingredient list of the chosen product from any shopping service/ingredient database and click on the “Classifier” link.
   OR
   Click on the “Sample ingredients” link and copy any ingredient list from the list provided and click on the “Go to Classifier” link below the table.

2. Paste the ingredient list in the form box and click classify to see how safe the product is ranked on a scale of 1-5 where 5 is the safest.

Advantages:

1. Helps in avoiding consumption of high amounts of harmful additives.
2. Provides an easy to use interface.

Disadvantages:

1. The tool classifies all low risk additives in class 5, which may not be helpful in scenarios like high sugar content.
2. Ingredient list has to be copy pasted or manually entered into the tool.

Future Scope:

1. More samples should be collected and more features should be fed to the model to cover more scenarios and increase efficiency.
2. A mobile application which scans ingredients using image to text conversion should be built using the existing API.

Code:

https://github.com/swathi0710/AIMLCCE_Final-Project