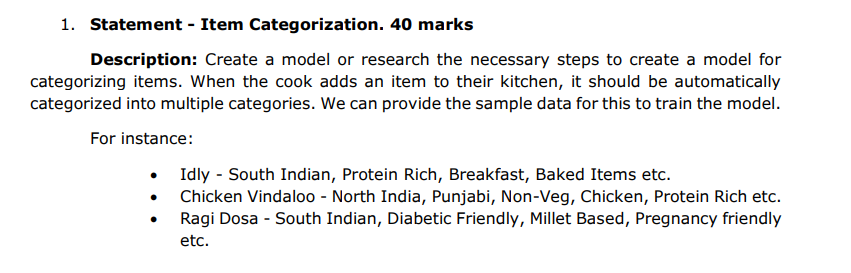
**Cookr Hackathon - Task 02**

**Team Members - Sahil Jamadar SK - 21PC27**

**- Harshad Krishna B - 21PC15**

**Problem Statement :**

**Approach 1: Model-Pipelining Classification**

**Objective:**

Predict the category of a dish (Meal\_Type, Cuisine\_Type, etc.) using a single classification model.

**Steps:**

**Data Preprocessing:**

Cleaned and preprocessed the dataset, handling missing values and encoding categorical variables.

Applied feature engineering, including vectorization of text data using TfidfVectorizer for the 'Ingredients' column.

**Model Training:**

Utilized a Random Forest Classifier for each target variable and pipelining the output of the models to predict another target variable (Meal\_Type, Cuisine\_Type, etc.).

Employed Label Encoding for categorical target variables by using OneHotEncoder from sklearn library

**Evaluation:**

Evaluated model performance using accuracy score.

Provided predictions for a few test cases.

**Approach 2: Multi-Target Regression (MTR) with Clustering**

**Overview:**

Approach 2 involves using Multi-Target Regression (MTR) with Clustering to categorize foods based on various attributes. This approach considers the relationships between different target variables, creating a more holistic model.

**Steps:**

**Data Preprocessing:**

**Text Data Processing:**

Extracted information from the 'Ingredients' column.

Applied TfidfVectorizer for converting text data into numerical format.

**Categorical Feature Encoding:**

Utilized One-Hot Encoding for categorical features such as 'Diet', 'Flavor', 'Carbohydrate\_Content', 'Protein'.

Ensured proper encoding of categorical features for model input.

**Clustering:**

Employed KMeans clustering to group similar data points.

'Ingredients' and one-hot encoded categorical features were considered for clustering.

**Model Training:**

**Target-Specific Regressors:**

For each cluster, trained Decision Tree Regressors for each target variable (Meal\_Type, Cuisine\_Type, Preparation\_Method, Health\_Benefits).

Considered all encoded features except the target variable and the 'Cluster' label.

**Prediction:**

Predicted each target variable for instances in the test set.

Utilized the trained regressors for accurate predictions.

**Evaluation:**

**Mean Squared Error:**

Calculated Mean Squared Error (MSE) for each target variable to assess the model's performance.

Evaluated the overall effectiveness of the approach.

Models and Mathematical Functions:

**Root Mean Squared Error (RMSE):**

Similar to MSE but takes the square root of the average squared differences.

RMSE is in the same unit as the target variable, making it more interpretable.

**R-squared (R2) Score:**

Measures the proportion of the variance in the dependent variable that is predictable from the independent variables.

Ranges from 0 to 1, with higher values indicating better model performance.

**Decision Tree Regressors:**

Used Decision Trees for regression to model the relationship between features and target variables.

The regression trees capture non-linear relationships within each cluster.

KMeans Clustering:

Utilized KMeans for grouping similar data points into clusters.

Improved the accuracy of regression models by considering data similarities.

**TfidfVectorizer:**

Applied TfidfVectorizer to convert 'Ingredients' text data into a numerical format.

Captured the importance of each term within the 'Ingredients' column.

**One-Hot Encoding:**

Encoded categorical features using One-Hot Encoding.

Ensured compatibility of categorical features with regression models.

**Proof of Working:**

The Mean Squared Error values for each target variable provide quantitative proof of the model's accuracy.

The trained models successfully predict target variables for new instances, showcasing the effectiveness of the approach.

Conclusion:

Approach 2 combines the power of regression models with clustering to create a robust solution for food categorization, capturing the nuances of each target variable and enhancing predictive accuracy.