Meal Classification and Recipe Recommendation

Rahul Maheshwari (MT19027)

Gaurav Lodhi (MT19063)

P. Akshay Kumar (MT19094)

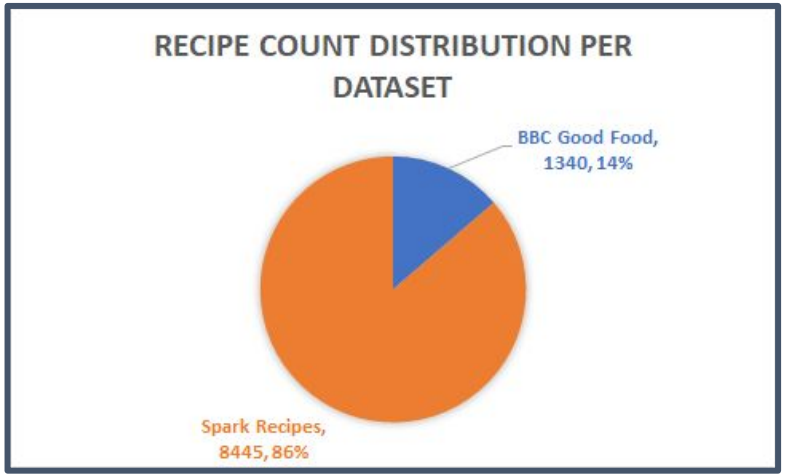
INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

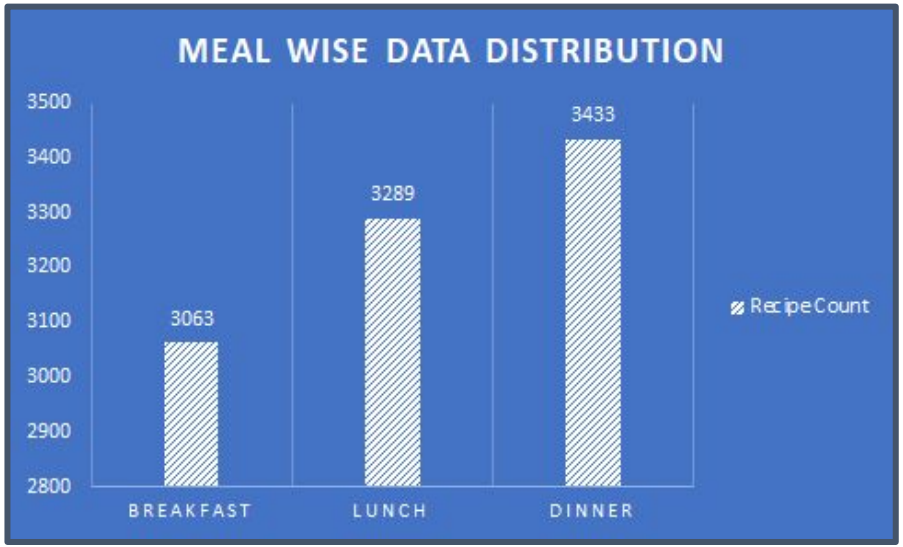
Tasks performed

- Scraping data from different sources and merging them with pre-processing.
- Classification models such as SVM, Decision Tree, Random Forest, KNN applied to predict the meal, given a set of ingredients.
- Recommending a meal to the user given a set of ingredients.
- Frequently co-occurring ingredients search for a given ingredient.
- Statistical plots such as frequency rank distribution and recipe size distribution for the dataset.
- GUI (Graphical User Interface) which integrates and binds all the above mentioned modules.

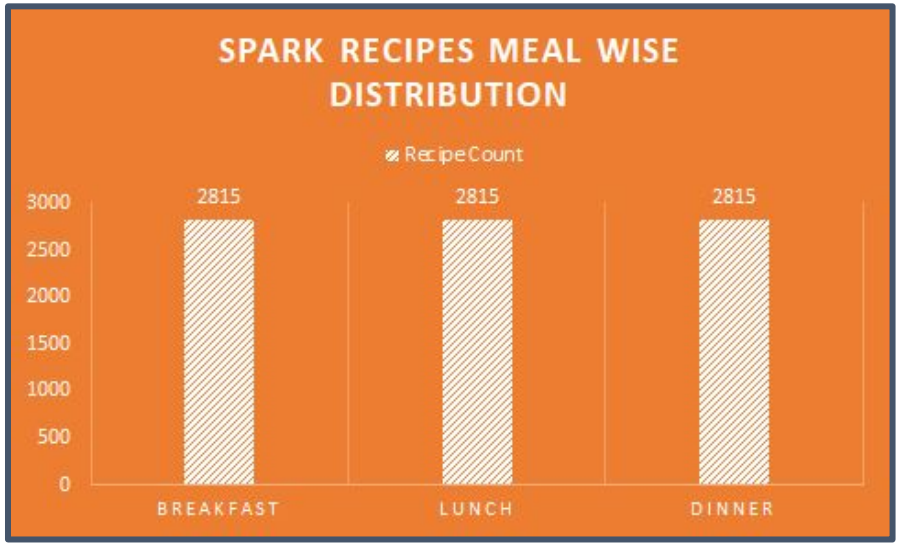
Food Data Sources and Distribution Plots

- BBC Good Food [https://www.bbcgoodfood.com/]
- Spark Recipes [https://recipes.sparkpeople.com/]





Meal wise distribution in datasets



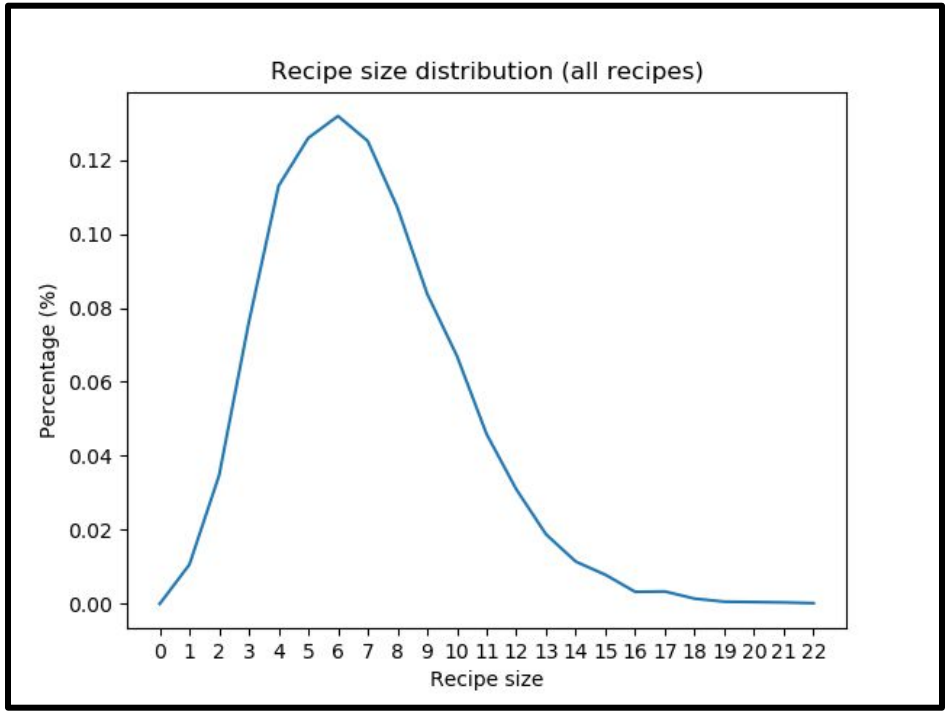
Features of the dataset

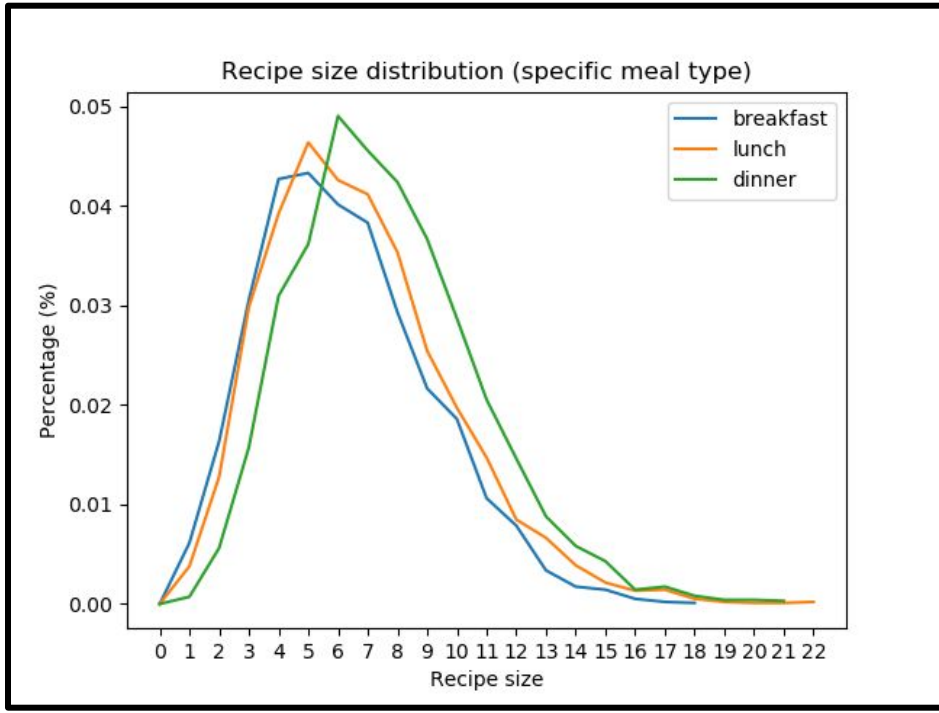
Recipe ID	Title	URL	Meal
Rating	Preparation Time	Cooking Time	Serves
Calories	Fat	Cholesterol	Carbs
Fiber	Protein	Ingredients	Lookup Ingredients
Rating Score	Cooking instructions		

Pre-processing

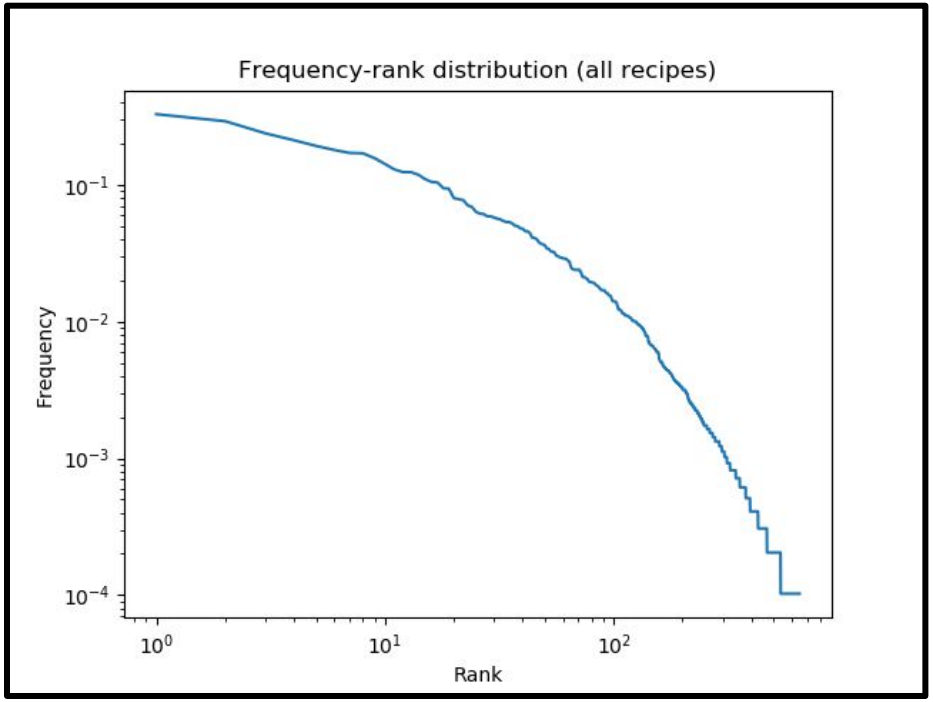
1. Data is scraped from multiple sources, hence all common features are merged and converted to same metric.
2. Ingredients are extracted from the scraped dataset and lemmatized.
3. Numerical features such as Ratings, Calories, Fat, Fiber, Protein, etc are normalized for unbiased classification and recommendation.
4. Some variation of preprocessing is performed separately for the required task at hand.

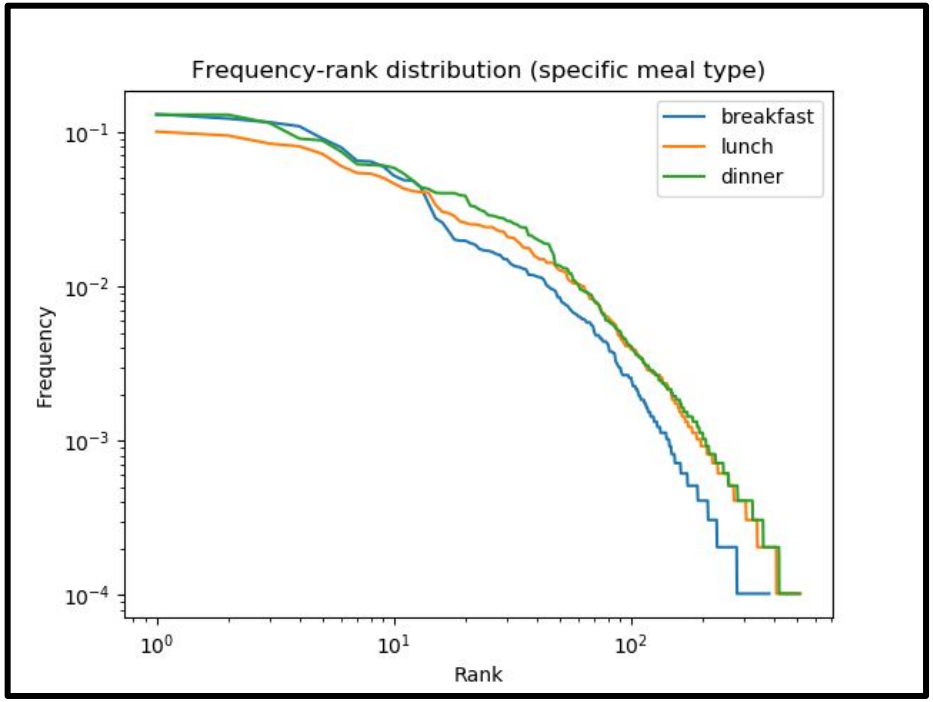
Recipe Size Distribution





Frequency Rank Distribution





Main screen GUI Screenshot

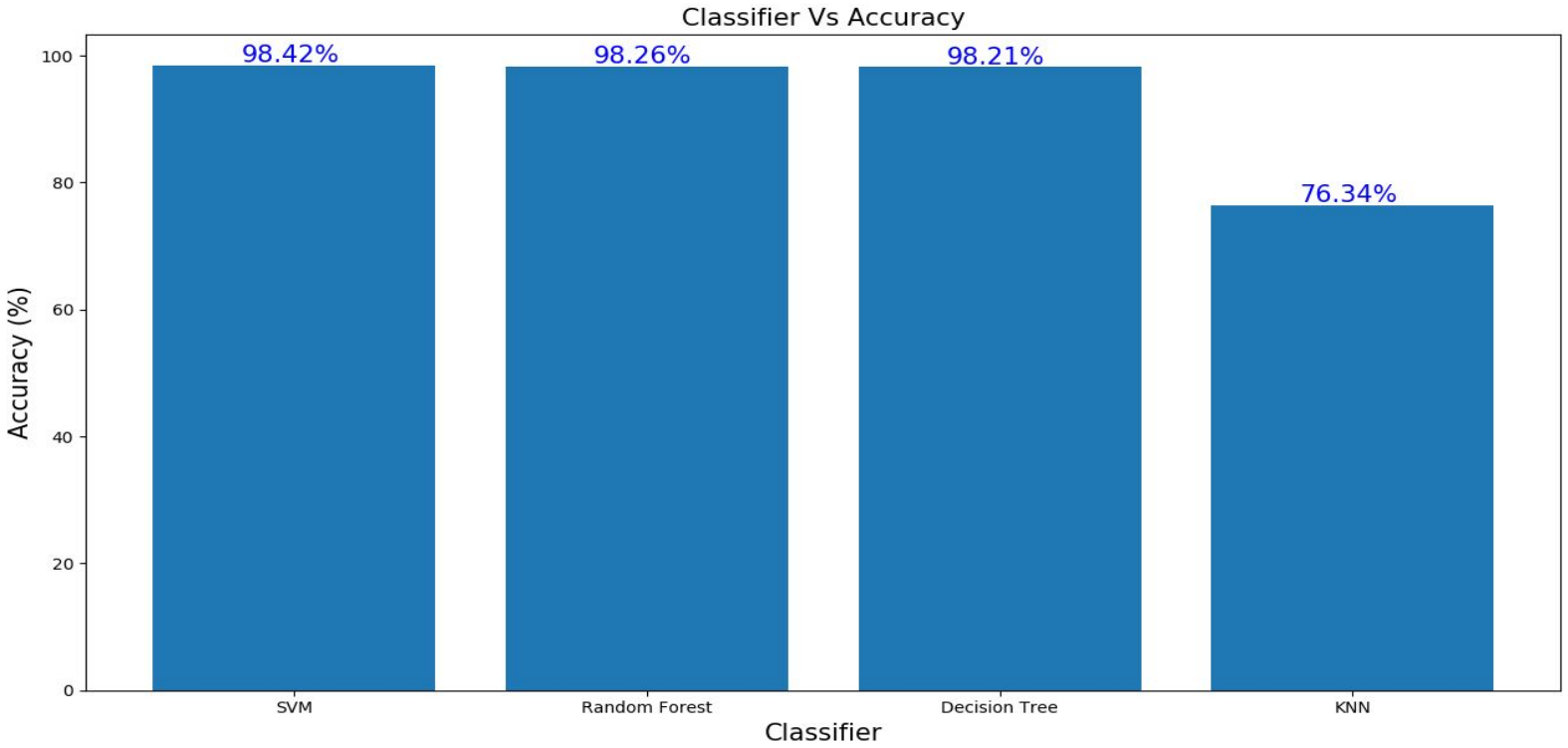
Task: Classification

- For classification of recipes in one of the meal category (Breakfast, Lunch, Dinner), several machine learning models are applied. Random Forest, Decision Tree, K Nearest Neighbor and Support Vector Machine are applied.

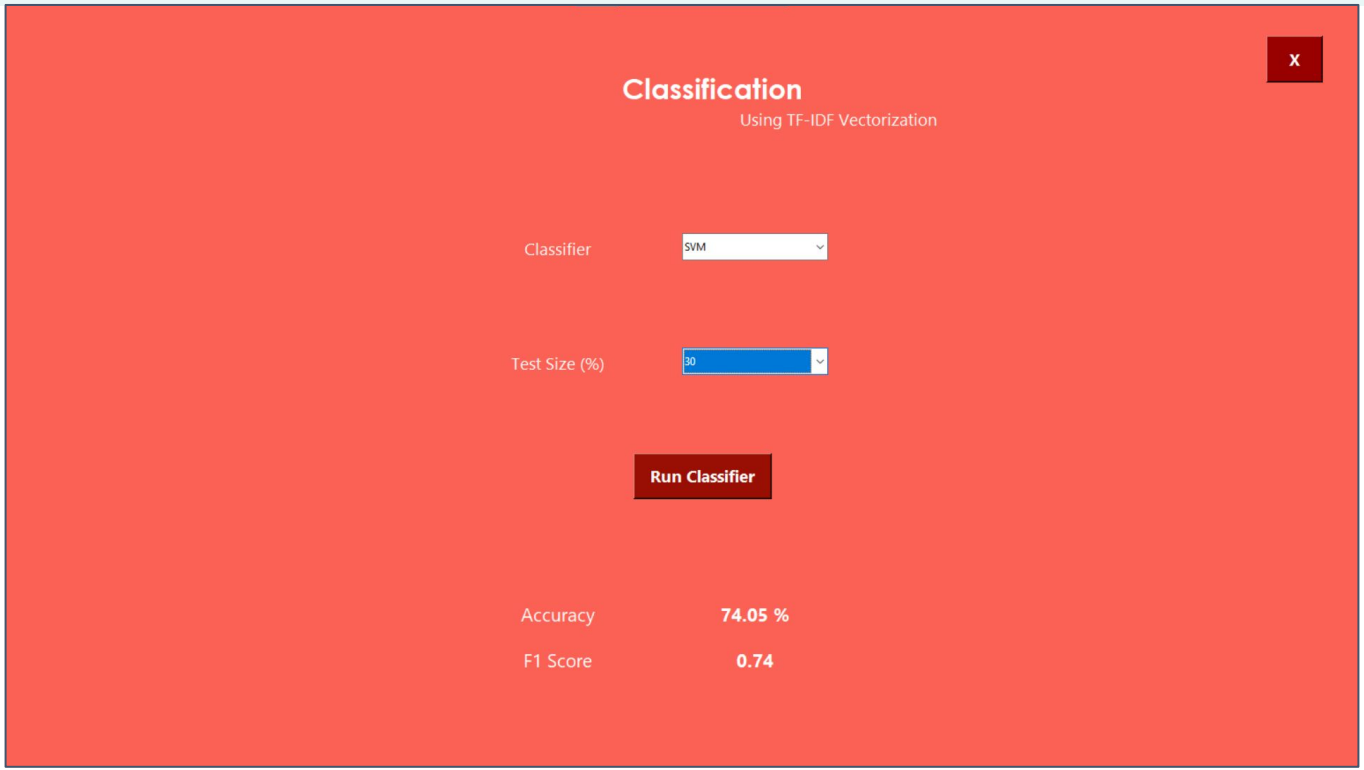
- GUI includes 2 screens, one for evaluation of the performance of various models by split the data into train and test and other for classifying a new recipe's meal category based on the ingredients.

- Best accuracy of 98.42 % was obtained with SVM which is used for classifying new recipes.

Classification Model performance

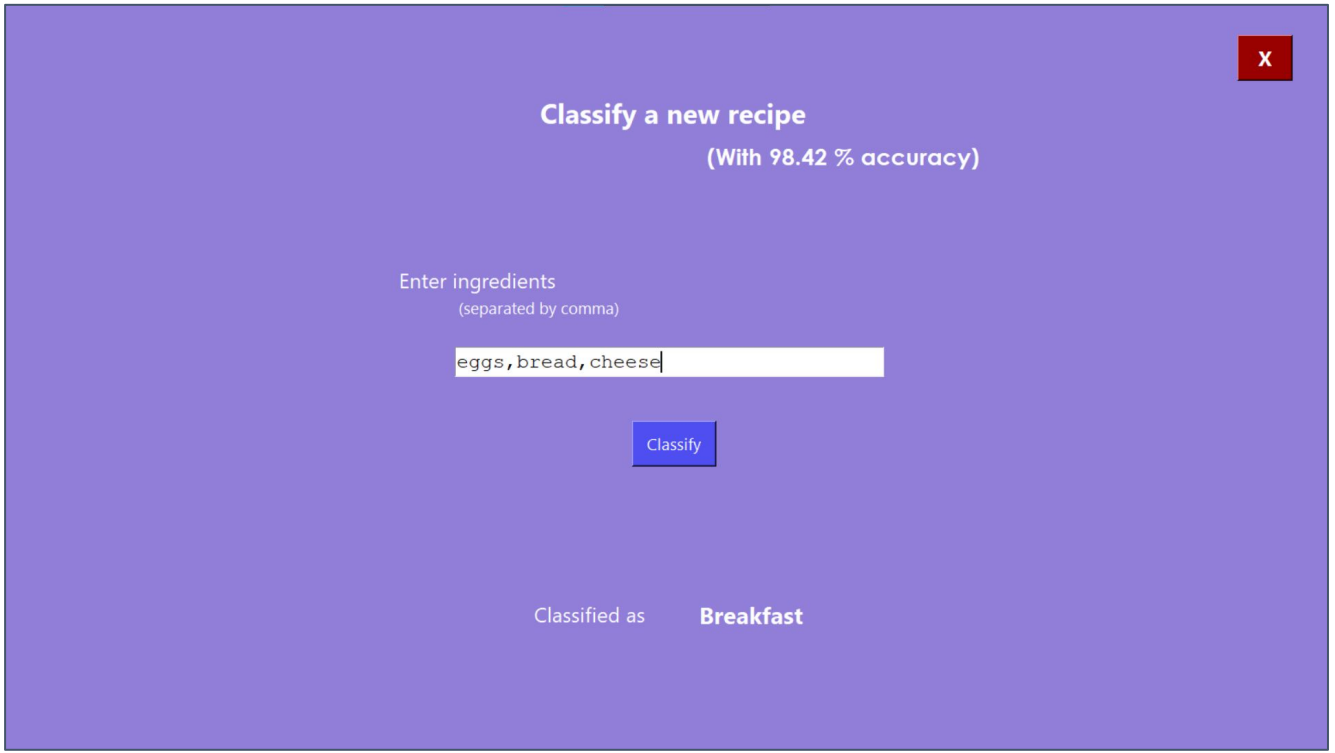


Classification Screenshot



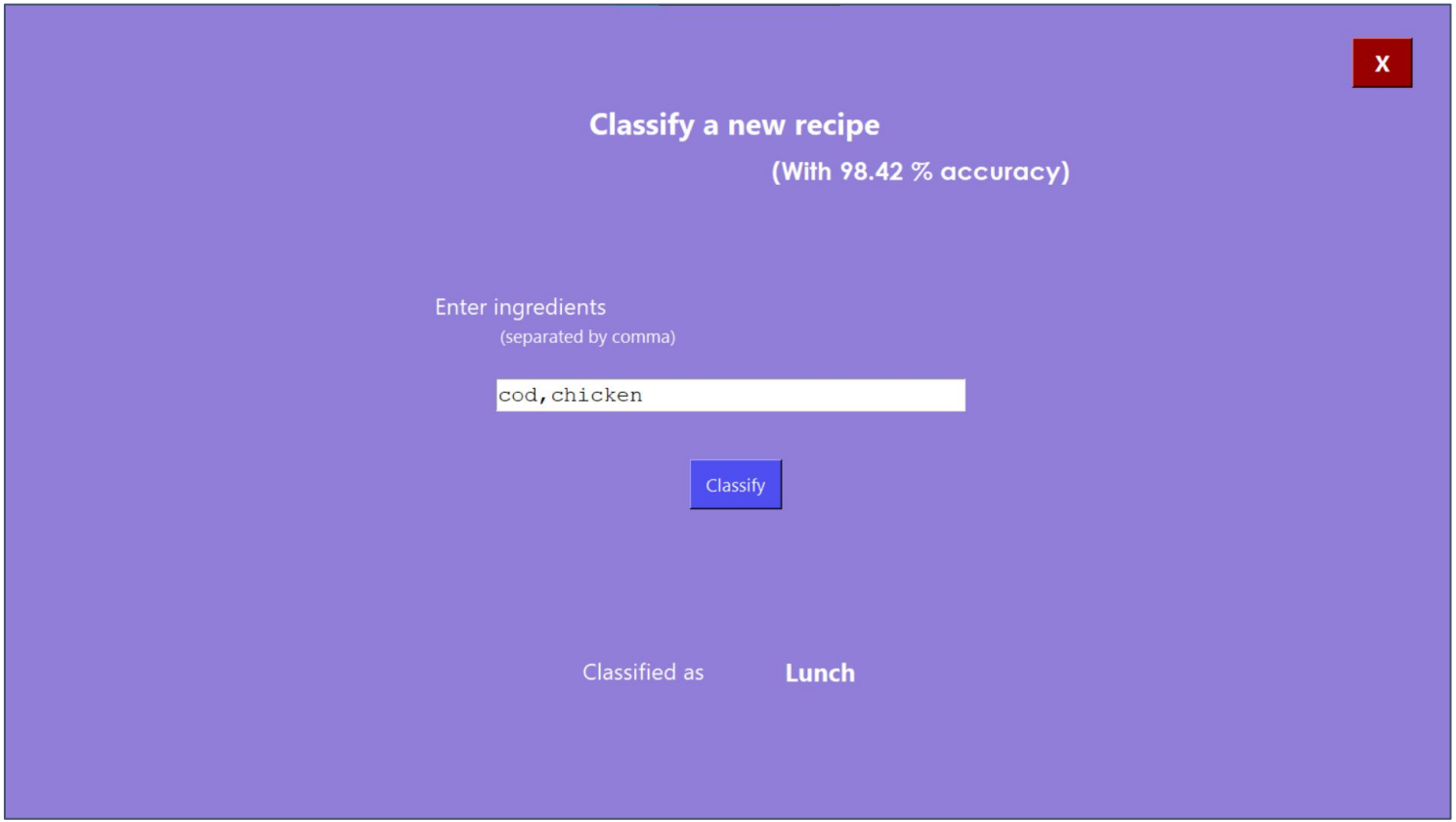
User can select any model and test size from given drop downs and evaluate the model performance.

Classify a New recipe screenshot – Breakfast

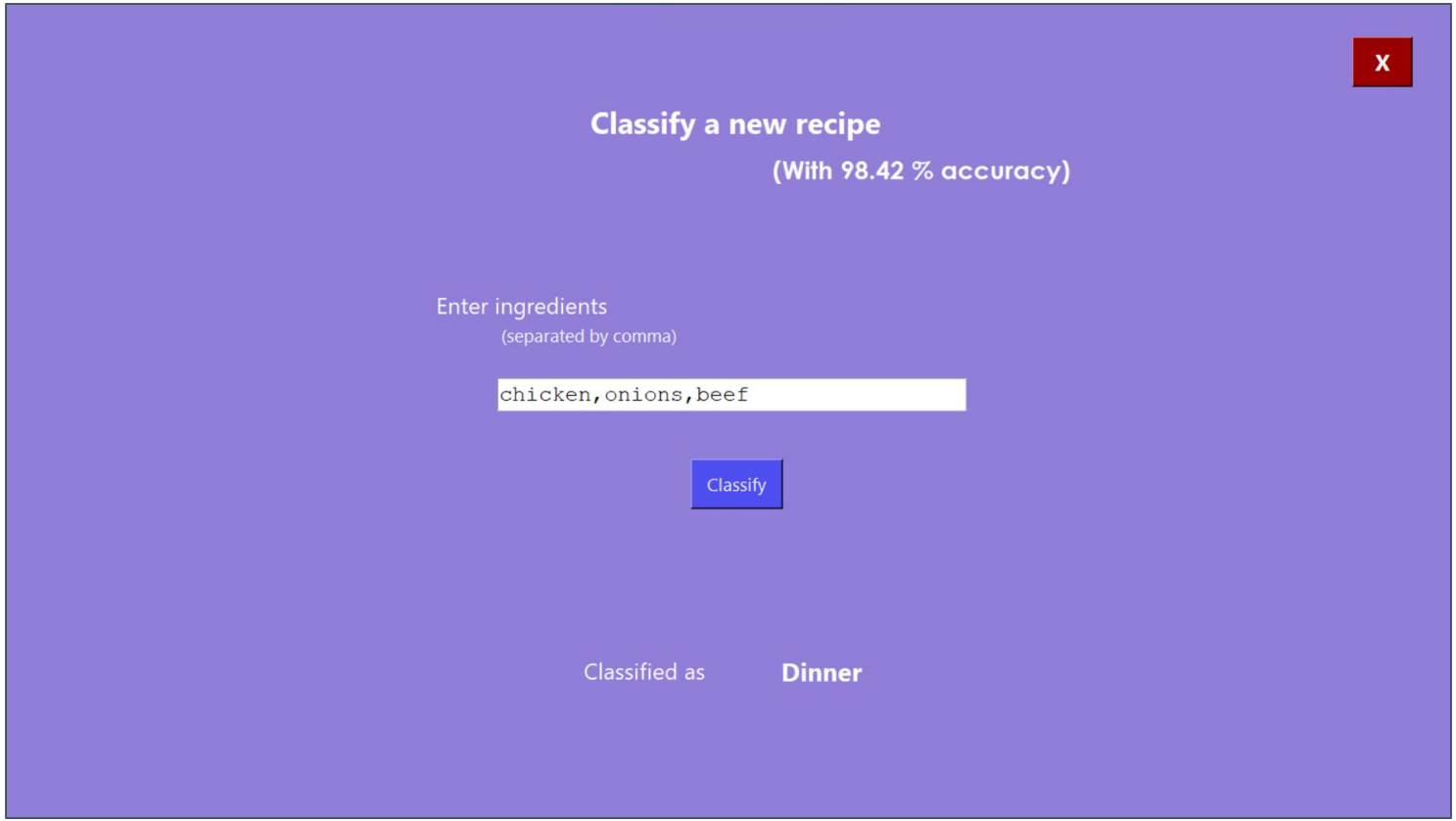


Input Ingredients are vectorized and meal type is predicted using pre-trained SVM model.

Classify a New recipe screenshot – Lunch



Classify a New recipe screenshot – Dinner



Recommending a meal

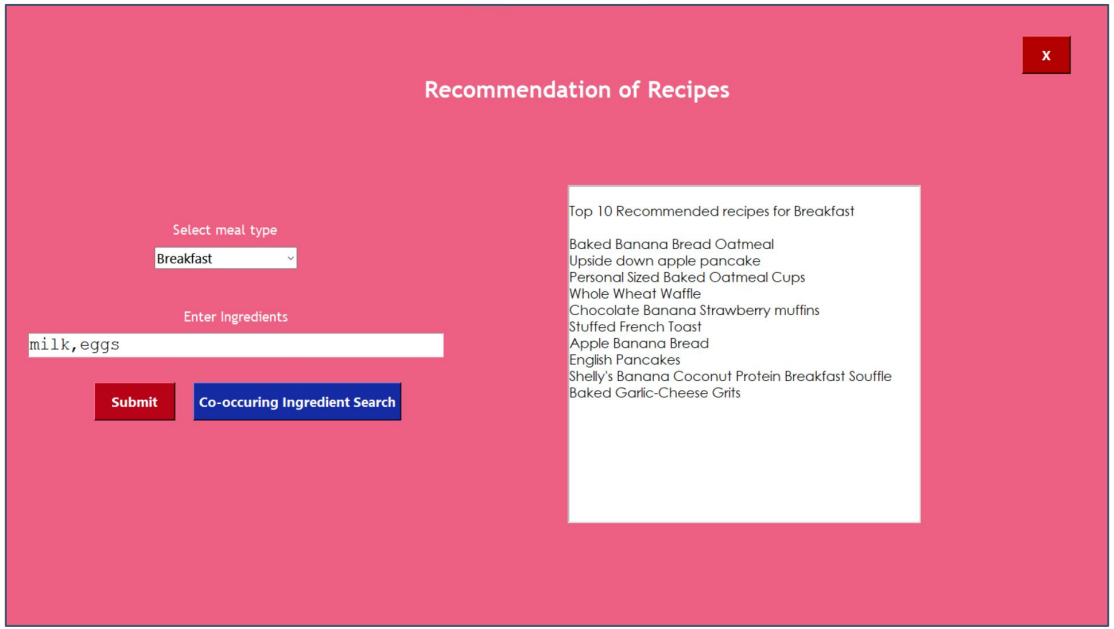
- For recommendation of the recipes for a specified meal type, recommendation system uses data features (Rating Score, Carbs, Fiber, Protein, Calories, Fat, Cholesterol) to calculate a recommendation score.

- Features were divided into two categories positive and negative category and tuned the model for the best recommendations for input ingredients.

- Based on the recommendation score, top 10 recipes for selected meal type and input ingredients are recommended by the system.

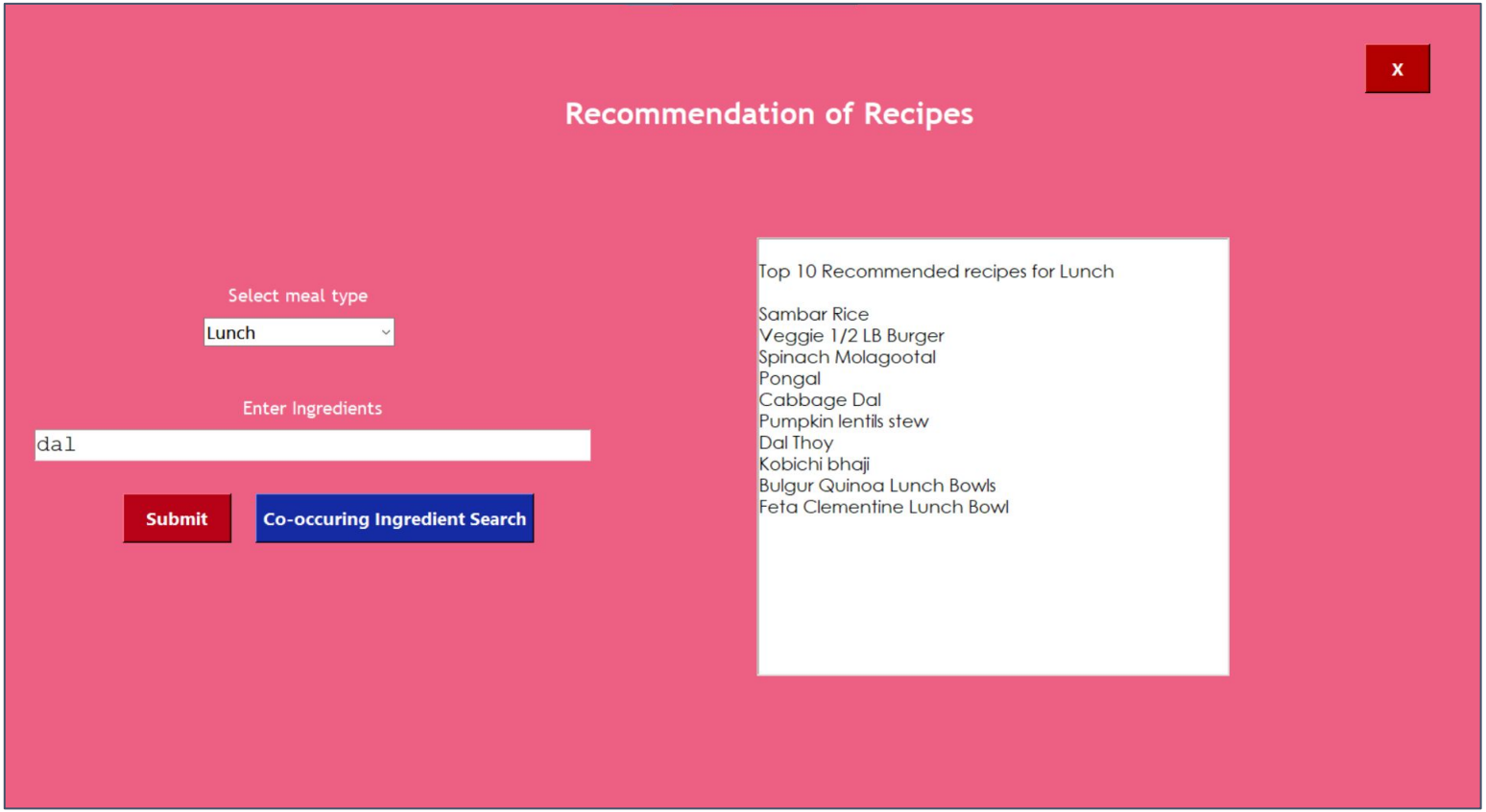
- We made our model flexible to adapt any changes if suggested by some fitness expert.

Recommendation Screenshot – Breakfast

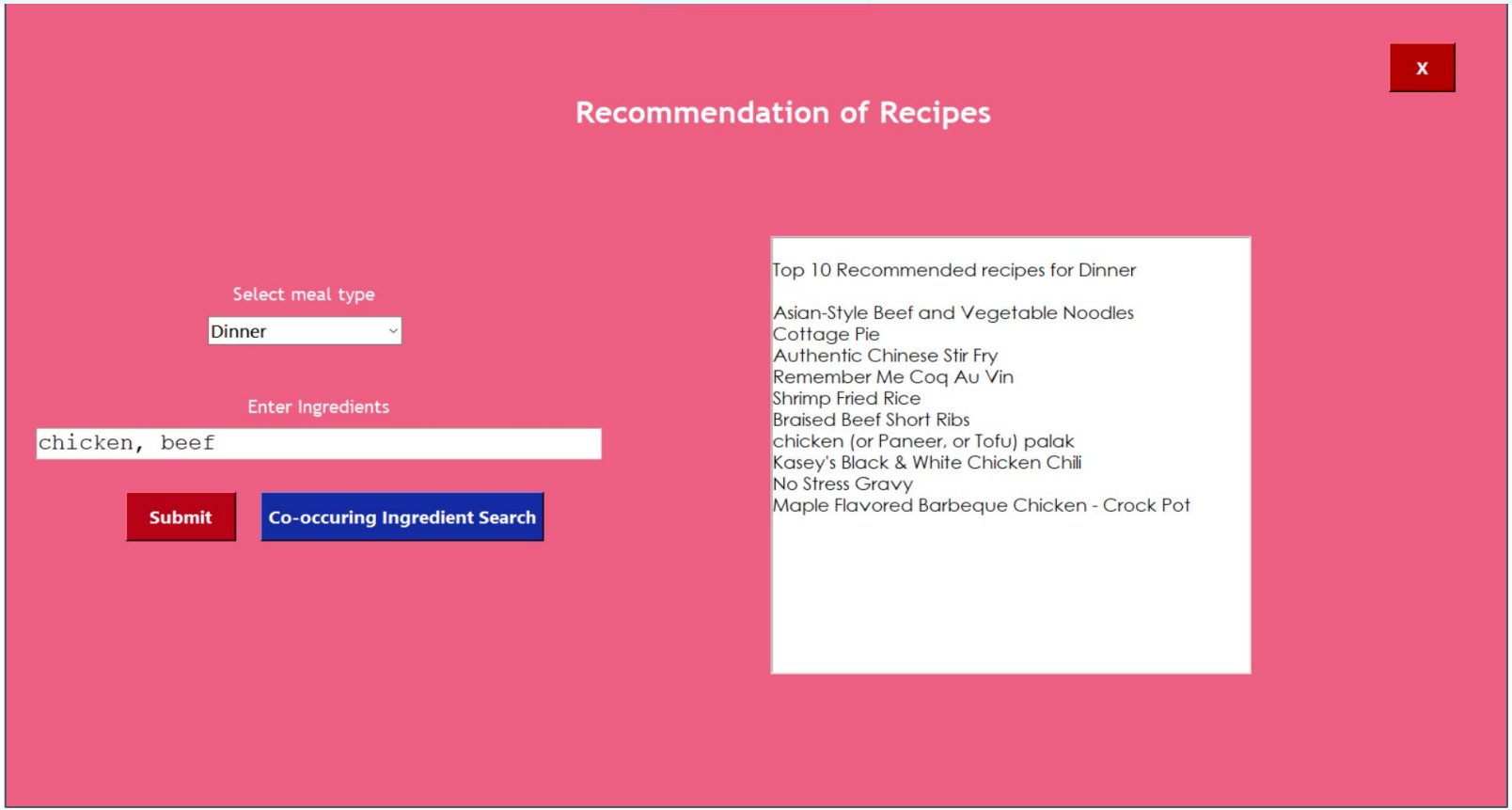


Based on selected meal type and input ingredients a recommendation score is calculated based on the corresponding recipe features and top 10 recipes are recommended.

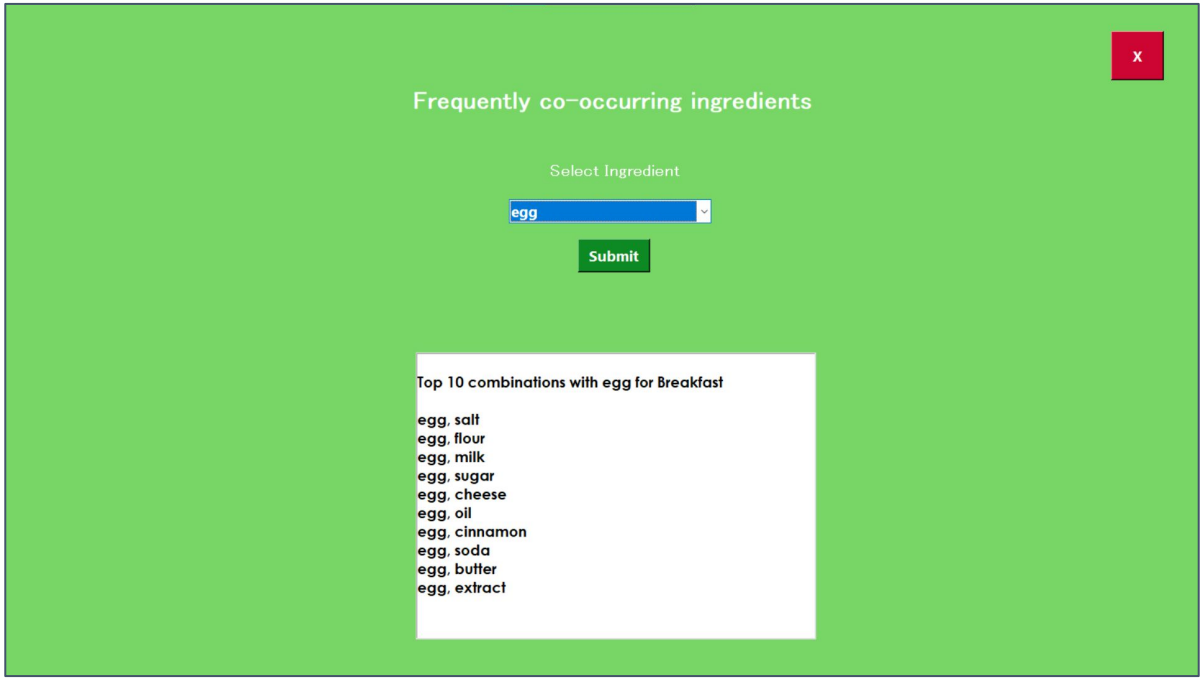
Recommendation Screenshot – Lunch



Recommendation Screenshot – Dinner

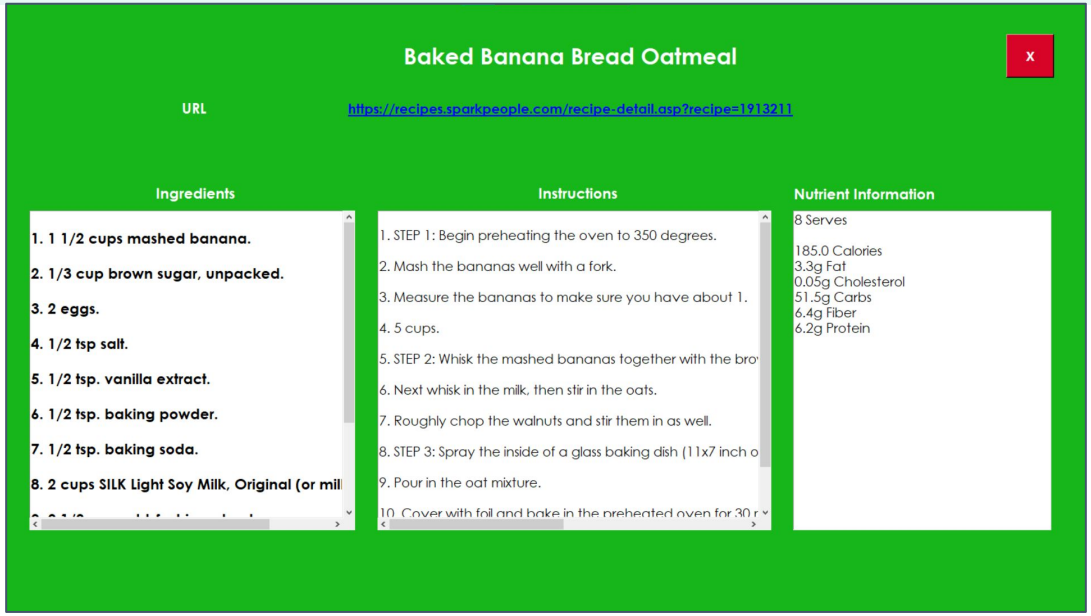


Additional Functionality
Co-occurring Ingredient Search



Based on meal type and ingredients entered in recommendation screen, frequently co-occurring ingredients are shown with the help of Apriori Algorithm frequent itemset mining.

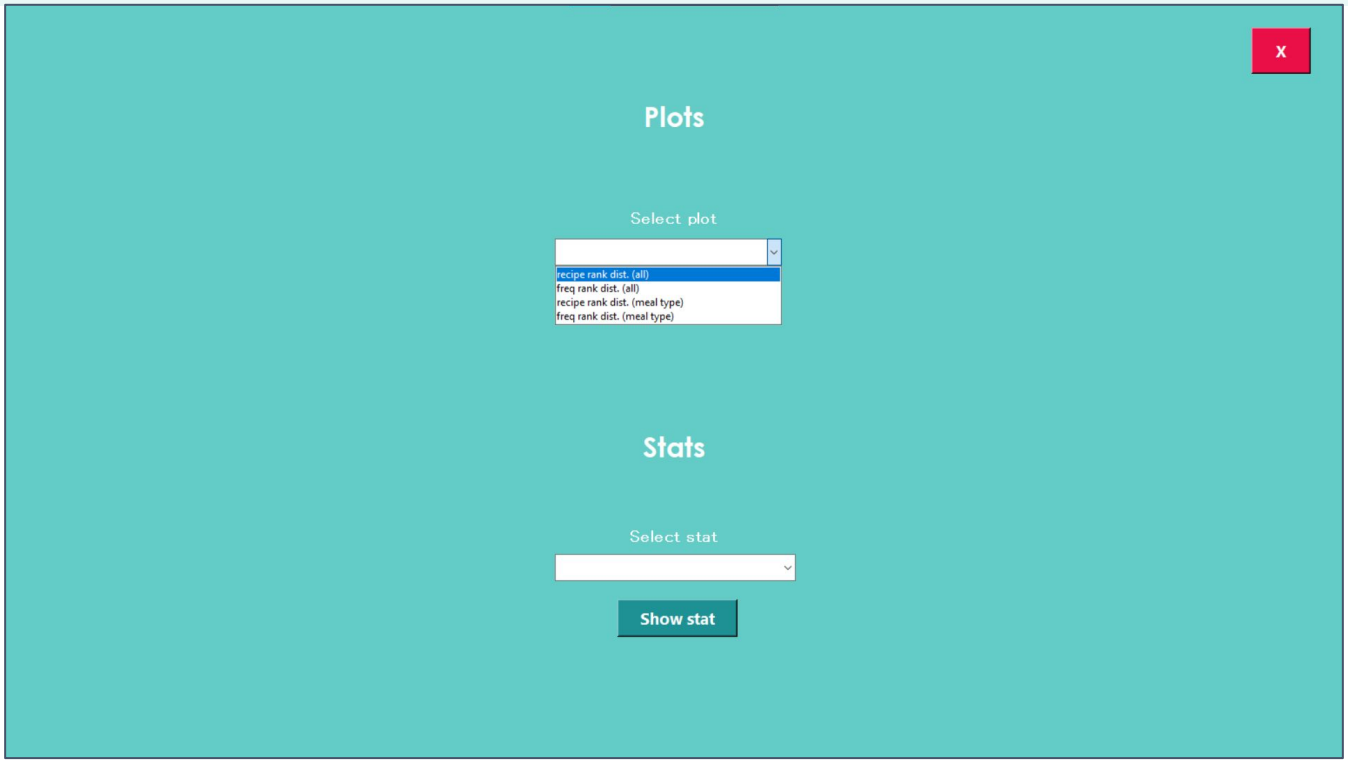
More information for a recommended recipe



More information about a recipe such as URL, complete list of ingredients, cooking instructions, etc. can be obtained by clicking on the recipe name.

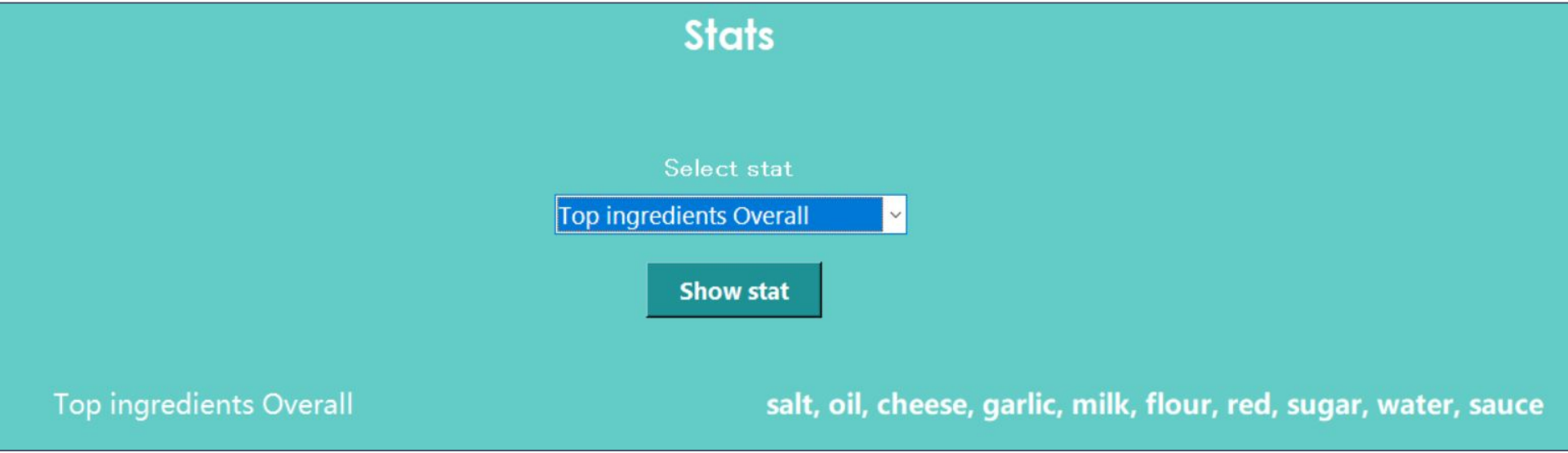
URL can be directly clicked to visit the recipe webpage for complete information and recipe images.

Plots and Stats Screenshot



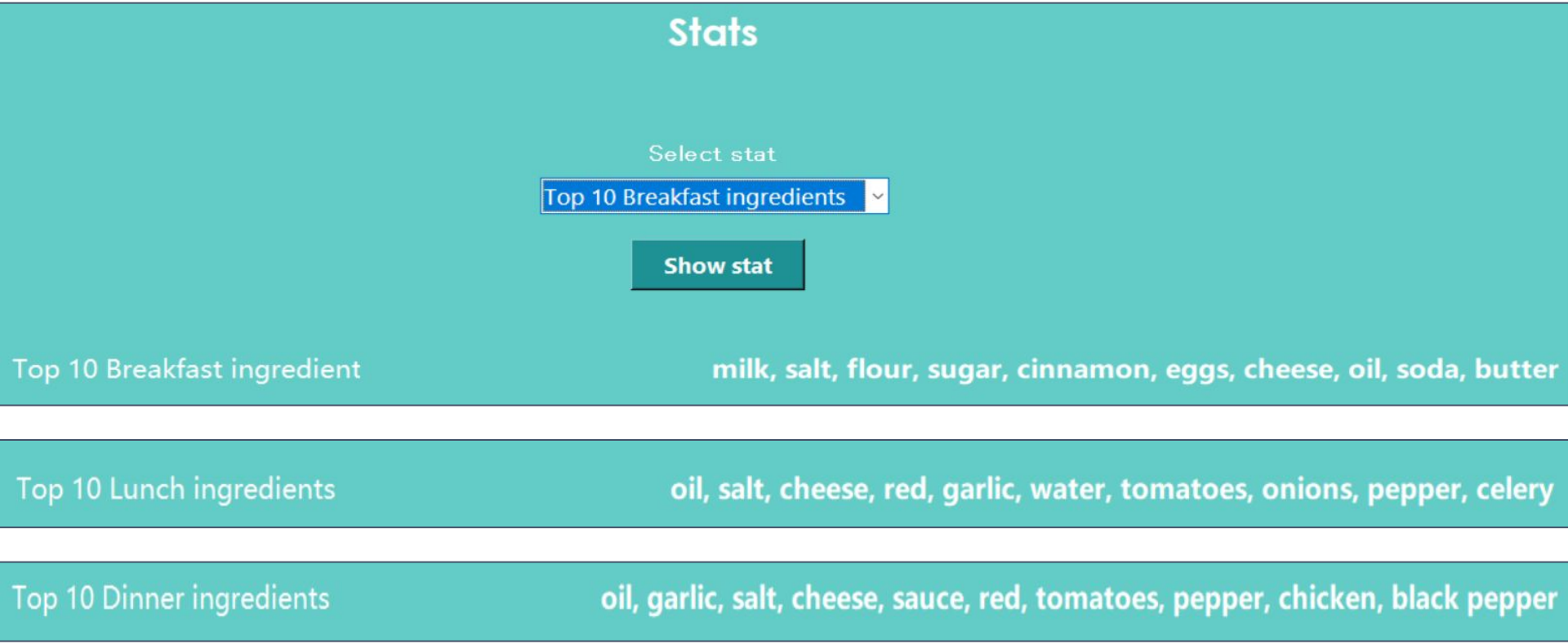
Various Plots for analyzing the dataset can be obtained using the dropdown menu for corresponding plot. All plots are mentioned in earlier slides.

Stats



Interesting stats from the dataset were calculated and displayed.

Stats (contd..)


Stats
Select stat
Top 10 Breakfast ingredients
Show stat
Top 10 Breakfast ingredient
milk, salt, flour, sugar, cinnamon, eggs, cheese, oil, soda, butter
Top 10 Lunch ingredients
oil, salt, cheese, red, garlic, water, tomatoes, onions, pepper, celery
Top 10 Dinner ingredients
oil, garlic, salt, cheese, sauce, red, tomatoes, pepper, chicken, black pepper

Contributions

- **Rahul Maheshwari** - Preprocessing, Classification, GUI.
- **P. Akshay Kumar** - Dataset scraping, Preprocessing, Plots & Stats, Co-occurring ingredient search.
- **Gaurav Lodhi** - Preprocessing and Recommendation.

Thank You!