

ИТОГОВЫЙ ПРОЕКТ

Создание прототипа
рекомендательного сервиса
рецептов

Гарипова Ляйсан Мухтаровна
Data Scientist (DS 102)
Haggopai (Uranus)



2022

РЕФЛЕКСИЯ

В данном проекте я узнала :

- Как подготовить и очистить датасет с большим количеством данных для дальнейшего обучения
- Использование нескольких алгоритмов с разными метриками и выбор лучшей модели
- как работать в PyCharm

Задание было понятно, но возникли трудности в длительных расчетах модели, работе с API и разработкой программы в PyCharm.

По завершению проекта я похвалила бы себя за усердие и терпение.

Цель работы

Закрепление навыков, полученных в ходе выполнения предыдущих задач.
Создание прототипа рекомендательного сервиса рецептов

Входные данные и инструменты

- датасет из Epicurious
- API FoodData
- сайт www.epicurious.com -pdf с нутриентами

Этапы подготовки и разработки

- | | |
|--------------------------|---------------------------|
| 1. Исследование | 2. Разработка |
| 1.1. Прогноз | 2.1. Класс Forecast |
| 1.1.1. Подготовка данных | 2.2. Класс NutritionFacts |
| 1.1.2. Регрессия | 2.3. Класс SimilarRecipes |
| 1.1.3. Классификация | 3. Основная программа |
| 1.1.4. Принятие решения | 4. Бонусная часть |
| 1.2. Пищевая ценность | |
| 1.3. Похожие рецепты | |

1.

**Подготовка
данных**

На входе был дан датасет `eri_r.csv` содержащий 680 столбцов и 20052 строки.

При помощи дополнительного json файла с ингредиентами из датасета были удалены не ингредиенты

После анализа столбцов, на предмет отображения ингредиента и удаления столбцов , которые не содержат ингредиенты, в датасете осталось 339 столбцов

2. РЕГРЕССИЯ

Модели

- LinearRegression {'fit_intercept': True}
- DecisionTreeRegressor {'max_depth': 15, 'max_features': 10, 'min_samples_split': 10}
- **RandomForestRegressor** {'max_depth': 50, 'max_features': 10, 'n_estimators': 100}

Ансамбли

- VotingRegressor {'weights': (1, 3, 2)}
- **BaggingRegressor** {'n_estimators': 5}

Метрика RMSE

Кросс-валидация на train

таргет rating 8 позиций

*точность наивной
модели Выше,
поэтому
отказываемся от
регрессии*

3. Классификация

Модели

- LogisticRegression {'fit_intercept': True, 'penalty': 'l1', 'solver': 'liblinear'}
- SVM {'kernel': 'rbf', 'C': 5}
- DecisionTreeClassifier {'class_weight': None, 'criterion': 'gini', 'max_depth': 2}
- **RandomForestClassifier** {'class_weight': None, 'criterion': 'gini', 'max_depth': 30, 'n_estimators': 150}

Метрика accuracy

Кросс-валидация на train

таргет rating 6 позиций

Смена таргета и метрики

*точность наивной
модели ниже*

3. Классификация

Модели

- LogisticRegression {'fit_intercept': True, 'penalty': 'l2', 'solver': 'liblinear'}
- SVM {'kernel': 'linear', 'C': 5}
- DecisionTreeClassifier {'class_weight': None, 'criterion': 'gini', 'max_depth': 2}
- RandomForestClassifier {'class_weight': None, 'criterion': 'entropy', 'max_depth': 30, 'n_estimators': 150}

Метрика accuracy

Кросс-валидация на train

таргет rating 6 позиций

Смена таргета

*точность наивной
модели ниже*

Классификация

Точностью (precision) называется доля правильных ответов модели в пределах класса —

это доля объектов действительно принадлежащих данному классу относительно всех объектов которые система отнесла к этому классу.

Что нам и требуется.

Модели

- LinearRegression {'fit_intercept': True, 'penalty': 'l1', 'solver': 'liblinear'}
- SVM {'kernel': 'linear', 'C': 5}
- DecisionTreeClassifier {'class_weight': 'balanced', 'criterion': 'entropy', 'max_depth': 1}
- **RandomForestClassifier** {'class_weight': None, 'criterion': 'gini', 'max_depth': 10, 'n_estimators': 150}

Ансамбли

- VotingClassifier {'weights': (1, 3, 2)}
- **BaggingClassifier** {'n_estimators': 5}

Метрика precision

Кросс-валидация на train

таргет rating 3 позиций

смена таргета

*точность наивной
модели ниже*

4. Принятие решения

Наилучшие результаты у модели :

Классификация

Таргет: 3 класса

Метрика: precision

Модель: RandomForestClassifier



5. Пищевая ценность и рецепты

Данные:

API FoodData

PDF с нутриентами

Рецепты:

www.epicurious.com



	Title	Rating	Almond	Amaretto	Anchovy	Anise	Apple	Apple Juice	Apricot	Artichoke	...	Wine	Wok	Yellow Squash	Yogurt	Yuca	Zucchini	URL	val	meal	class
0	Lentil, Apple, and Turkey Wrap	2.500	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Lentil%2C%20...	12.0	other	so-so
1	Boudin Blanc Terrine with Red Onion Confit	4.375	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Boudin%2C%20...	10.0	other	great
2	Potato and Fennel Soup Hodge	3.750	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Potato%2C%20...	6.0	other	great
3	Mahi-Mahi in Tomato Olive Sauce	5.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Mahi%2C%20Ma...	6.0	dinner	great
4	Spinach Noodle Casserole	3.125	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Spinach%2C%2...	6.0	other	so-so
...
27899	Chinese Barbecued Spareribs	3.750	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Chinese%2C%2...	8.0	other	great
27900	Artichoke and Parmesan Risotto	4.375	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Artichoke%2C...	4.0	dinner	great
27901	Turkey Cream Puff Pie	4.375	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Turkey%2C%20...	6.0	dinner	great
27902	Snapper on Angel Hair with Citrus Cream	4.375	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Snapper%2C%2...	12.0	other	great
27903	Baked Ham with Marmalade-Horseradish Glaze	4.375	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	https://www.epicurious.com/search/Baked%2C%20H...	6.0	other	great

27904 rows × 343 columns

ПРИМЕНЕНИЕ ПРОГРАММЫ

```
C:\Users\09\PycharmProjects\pythonProject1\venv\Scripts\python.exe C:\Users\09\PycharmProjects\pythonProject1\data\nutritionist.  
Пример ввода продуктов: Rice, Tomato, Beef  
Введите список ингредиентов на английском с большой буквы через запятую: Lemon, Chicken, Cheese
```

I. НАВ ПРОГНОЗ

great. Хороший выбор продуктов!

II. ПИЩЕВАЯ ЦЕННОСТЬ

Lemon

Total Lipid (fat) - 38.0 % of Daily Value
Sodium - 55.0 % of Daily Value
Fatty acids - 15.0 % of Daily Value
Vitamin C - 4.0 % of Daily Value
Carbohydrate - 2.0 % of Daily Value

Chicken

Protein - 48.0 % of Daily Value
Cholesterol - 27.0 % of Daily Value

Fatty acids - 18.0 % of Daily Value

Cheese

Fatty acids - 80.0 % of Daily Value
Total Lipid (fat) - 42.0 % of Daily Value
Calcium - 39.0 % of Daily Value
Cholesterol - 33.0 % of Daily Value
Protein - 32.0 % of Daily Value
Sodium - 26.0 % of Daily Value
Potassium - 2.0 % of Daily Value
Carbohydrate - 1.0 % of Daily Value

III. ТОП-3 ПОХОЖИХ РЕЦЕПТА:

- Grilled Lemon-Honey Chicken Breasts , peBtMcr: 3.125 , URL: <https://www.epicurious.com/search/lemon>
- Marinated Chicken Kabobs with Lemon Popper Yogurt Sauce , peBtMcr: 4.375 , URL: <http://www.epicurious.com>
- Slow-Cooker Moroccan Chicken Stew With Preserved Lemon and Olives , peBtMcr: 9.0 , URL: <https://www.epicurious.com>

Предложите мне на день Yes / No : [Yes](#)

МЕНЮ НА ДЕНЬ

ЗАВТРАК

Sour Cream-Streusel Coffee Cake (peBtMcr: 4.375)

Sour Cream-Streusel Coffee Cake (peBtMcr: 4.375)

Ингредиенты:

- Cake
- Chocolate
- Cinnamon
- Coffee
- Currant
- Sour Cream
- Walnut

Nutrients:

- Fatty acids : 275.0 %
- Total lipid (fat) : 150.0 %
- Sugars : 84.0 %

- Iron : 67.0 %
- Phosphorus : 63.0 %
- Vitamin C : 61.0 %
- Vitamin K (phylloquinone) : 47.0 %
- Zinc : 45.0 %
- Manganese : 44.0 %
- Selenium : 36.0 %
- Potassium : 32.0 %
- Fiber : 29.0 %
- Niacin : 19.0 %
- Calcium : 19.0 %
- Folate : 16.0 %
- Sugars : 14.0 %
- Vitamin A : 6.0 %
- Cholesterol : 6.0 %
- Choline : 5.0 %
- Thiamin : 0.0 %
- Riboflavin : 0.0 %
- Vitamin B-12 : 0.0 %
- Vitamin B-6 : 0.0 %

- Phosphorus : 0.0 %
- Niacin : 6.0 %
- Vitamin C : 4.0 %
- Folate : 2.0 %
- Vitamin K (phylloquinone) : 2.0 %
- Choline : 2.0 %
- Vitamin B-6 : 0.0 %
- Vitamin E (alpha-tocopherol) : 0.0 %
- Vitamin D (D2 + D3) : 0.0 %
- Riboflavin : 0.0 %
- Vitamin B-12 : 0.0 %
- Vitamin A : 0.0 %
- Thiamin : 0.0 %
- Selenium : 0.0 %
- Copper : 0.0 %
- Zinc : 0.0 %

URL: <https://www.epicurious.com/search/Blue%20Corn%20Bread%20Bread>

DEER

Blue Corn Bread (peBtMcr: 3.75)

Ингредиенты:

- Bread
- Cornmeal

- Cornmeal

Nutrients:

- Carbohydrate : 48.0 %
- Protein : 24.0 %
- Fatty acids : 15.0 %
- Fiber : 11.0 %
- Iron : 11.0 %
- Sodium : 11.0 %
- Total lipid (fat) : 9.0 %
- Calcium : 2.0 %
- Vitamin C : 1.0 %
- Cholesterol : 0.0 %

URL: <https://www.epicurious.com/search/Blue%20Corn%20Bread>

ПРОВЕРКА ПРОЕКТА

DashboardCalendarProgressProjectsActivitiesMore

Peer Reviews

alainama

11 November 2022, 10:33

Отличная работа!
https://drive.google.com/file/d/1UJYpiEFcg_PRf-jkjmBAi5FVjk8rruPk/view?usp=sharing
<https://drive.google.com/file/d/1NIFOWxAhRqDhZeXKtDfigEMgCkjH9nND/view?usp=sharing>

MY SCORE

Спасибо за проверку!

4/4

4/4

4/4

4/4

cblossom_support

Today, 12:49

Хорошая работа, улучшить понимание работы моделей и будет вообще супер!
https://drive.google.com/file/d/1kl6rMWVTuiK_WAMas6_UgsGmitk87-Sg/view?usp=sharing

MY SCORE

Add feedback

Completed (120 XP, 120%)

Project is completed successfully

Try again

haggopai

Profile

Notifications

Change password

Wallet

Switch theme

Log out