

Иерархическая кластеризация для промышленности

Приложение загружает табличные данные, выполняет предобработку (очистка, удаление выбросов, кодирование категориальных и стандартизация числовых), а затем строит иерархическую кластеризацию. Есть визуализации (дендrogramма, 2D-проекция) и анализ признаков (профили кластеров и важности).

Требования к данным:

- Форматы: CSV, XLS, XLSX, XLSM
- Структура: первая строка - заголовки, строки - объекты, столбцы - признаки.
- Десятичный разделитель в CSV: запятая (,). Если в ваших данных точка, сохраните файл как XLSX или преобразуйте разделитель.

Инструкция по использованию приложения

1. Откройте приложение.
2. В блоке «Загрузите файл» перетащите CSV/XLS(X/M) или нажмите **Browse files**.



Иерархическая кластеризация для промышленности

Загрузите файл

Drag and drop file here
Limit 200MB per file • CSV, XLS, XLSX, XLSM

📁data.xls_with_data.xls 229.0KBBrowse files

File type • CSV, Excel (.csv, .xls, .xlsx, .xlsm)

При успешной загрузке появится зелёное сообщение с размером таблицы, пред просмотр данных и таблица типов.

Иерархическая кластеризация для промышленности

Загрузите файл

Drag and drop file here
Limit 200MB per file • CSV, XLS, XLSX, XLSM

📁data.xls_with_data.xls 229.0KBBrowse files

File type • CSV, Excel (.csv, .xls, .xlsx, .xlsm)

Загружено: data.xls_with_data.xls [1162x12]

3. Пред просмотр и типы столбцов

В секции «Пред просмотр данных» показаны первые строки.

В секции «Типы данных в столбцах» видно, как pandas распознал типы (int, float, object и т.п.).

Иерархическая кластеризация для промышленности

Загрузите файл

Drag and drop file here
Limit 200MB per file • CSV, XLS, XLSX, XLSM

Browse files

data_xls_with_data.xls 229.0KB

File type • CSV, Excel (.csv, .xls, .xlsx, .xslm)

Загружено: data_xls_with_data.xls [1162×12]

Предпросмотр данных

Номер Z-конфигурации	КлчПост/Потребн	Базовая ЕИ	Количество (Норматив)	VID_PROD_GP	TOL	SHIRINA	DLINA	VID_TO	MARKA	ST_PROD	Действие
0	401,077,983	53.526 Т	1,197.6 Р0	58	2,440	15,578	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
1	401,090,004	55.476 Т	1,197.6 Р0	32.2	2,260	15,578	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
2	401,090,006	47.301 Т	1,197.6 Р0	33	1,880	15,578	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
3	401,090,005	46.477 Т	1,197.6 Р0	38	1,609	15,578	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
4	401,089,979	143.408 Т	1,197.6 Р0	58	2,450	15,588	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
5	401,089,980	143.408 Т	1,197.6 Р0	58	2,450	15,588	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
6	401,089,988	64.591 Т	1,197.6 Р0	38	2,240	15,578	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
7	401,089,996	155.545 Т	1,197.6 Р0	54	1,750	15,588	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
8	401,089,997	78.05 Т	1,197.6 Р0	31.3	2,791	15,649	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	
9	401,089,987	48.273 Т	1,197.6 Р0	24.8	2,540	15,650	БЕЗ ТО	S355NL	EN 10025-3:2019(E)	2	

Типы данных в столбцах

Столбец	Тип данных
Номер Z-конфигурации	int64
КлчПост/Потребн	float64
Базовая ЕИ	object
Количество (Норматив)	float64
VID_PROD_GP	object
TOL	float64
SHIRINA	int64
DLINA	int64
VID_TO	object
MARKA	object
ST_PROD	object
Data	object

Далее нам будет доступен раздел “Предобработка данных”

4. Предобработка данных

■ Переименование столбцов

В блоке «Переименование столбцов» выберите столбец.

Ведите Новое имя и нажмите Переименовать. Если имя уже существует - получите предупреждение. После переименования интерфейс обновится.

■ Фильтр по дате

В «Фильтр по дате» выберите столбец с датой. Приложение попытается распарсить даты.

Если распознавание удалось, задайте диапазон в «Диапазон дат» и нажмите Отфильтровать. Данные и счётчики строк/столбцов обновятся.

■ Выбор признаков для кластеризации

В блоке «Признаки для кластеризации» исключите столбцы, которые использовать не хотите, и добавьте те, которые планируете использовать.

Предобработка данных

Переименование столбцов

Выберите столбец Новое имя

Номер Z-конфигурации Введите новое имя столбца Переименовать

Фильтр по дате

Выберите столбец с датой Диапазон дат

Номер Z-конфигурации 1970/01/01 – 1970/01/01 Отфильтровать

Признаки для кластеризации

Выберите столбцы

Номер Z-конфигурации, КлчПост/Потребн, Базовая ЕИ, Количество (Норматив), VID_PROD_GP, TOL, SHIRINA, DLINA, VID_TO, MARKA, ST_PROD, Data

■ Предобработка данных

Нажмите «Препроцессинг». После успеха внизу появится «Превью датасета» - это уже подготовленная матрица признаков для кластеризации.

■ Превью датасета

У вас будет возможность выбрать «Сколько строк показать» - выберите число строк предпросмотра. Под таблицей отображается «Размер датасета».

Препроцессинг данных (предобработка данных → IQR → encode/scale)

Препроцессинг														
Предобработка данных пройдена														
IQR по числовым признакам реализован														
Категориальные признаки закодированы, количественные признаки стандартизированы														
Превью датасета														
Сколько строк показать														
15														
1	ord__Базовая ЕИ	ord__VID_PROD_GP	ord__VID_TO	ord__MARKA	ord__ST_PROD	ord__Data	num__Номер Z-конфигурации	num__КвнПост/Потребн	num__Количество (Норматив)	num__TOL	num__SHIRINA	num__DLINA		
2	0	0	0	25	3	2	-2.759	0.4744	0.6465	0.169	0.3843	1.7542		
3	0	0	0	25	3	3	-2.759	0.2527	0.6465	0.2095	-0.3446	1.7542		
4	0	0	0	25	3	4	-2.7598	2.8597	0.6465	0.4628	-0.8644	1.7542		
5	0	0	0	25	3	5	-2.7598	2.8597	0.6465	1.4761	0.7487	1.7569		
6	0	0	0	25	3	6	-2.7595	0.7217	0.6465	0.4628	0.3459	1.7542		
7	0	0	0	25	3	7	-2.7593	3.1889	0.6465	1.2734	-0.5939	1.7569		
8	0	0	0	25	3	8	-2.7593	1.0868	0.6465	0.1234	1.4028	1.7732		
9	0	0	0	25	3	9	-2.7596	0.279	0.6465	-0.2059	0.9213	1.7735		
10	0	0	0	25	3	10	-2.7592	0.32	0.6465	-0.1755	0.9213	1.7735		
11	0	0	0	25	3	11	-2.7589	0.811	0.6465	-0.1451	0.9213	1.7735		

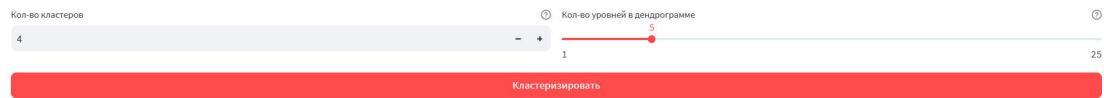
5. Иерархическая кластеризация

Доступна, когда предобработанные данные готовы и строк ≥ 2 .

Параметры:

- Кол-во кластеров - k (целое ≥ 2 ; рекомендовано начать с 4, затем попробовать 8 для детализации).
- Кол-во уровней в дендрограмме - глубина усечённой дендрограммы.

Иерархическая кластеризация



Нажмите «Кластеризовать». Появятся две вкладки:

- Визуализация результатов кластеризации

Дендрограмма

2D-визуализация - проекция кластеров функцией

■ Влияние признаков на кластеризацию

Тепловая карта средних значений признаков по кластерам - профиль каждого кластера.

Важность признаков при кластеризации - ориентировочные веса признаков (по логистической регрессии).

