Выполните задания, используя jupyter

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Датасет** | **Описание** | **Целевой признак** |
| 1 | <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud> | Необходимо, чтобы компании, предоставляющие кредитные карты, могли распознавать мошеннические операции по кредитным картам, чтобы с клиентов не взималась плата за товары, которые они не приобретали.  Датасет содержит операции, совершенные по кредитным картам в сентябре 2013 года европейскими держателями карт. | Class |
| 2 | <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset> | Датасет для определения риска сердечного приступа.  Необходимо проанализировать и предсказать, подвержен ли человек сердечному приступу или нет. | **Output** |
| 3 | <https://www.kaggle.com/datasets/shivam2503/diamonds> | Этот классический датасет содержит цены и другую информацию о почти 54 000 бриллиантов.  Проанализируйте бриллианты по их огранке, цвету, чистоте и научите модель предсказывать цену. | **price** |
| 4 | <https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset> | Система наблюдения за поведенческими факторами риска (BRFSS) - это ведущая в стране система телефонных опросов по вопросам здравоохранения, касающиеся их поведения, связанного с риском для здоровья, хронических заболеваний и использования профилактических услуг.  Создайте модель, которая будет использована в медицинской области для определения общего состояния здоровья пациента. | General Health |
| 5 | <https://www.kaggle.com/datasets/juhibhojani/house-price> | Датасет содержит информацию о недвижимости, включающую в себя разные характеристики о жилье.  Этот датасет дает представление о динамике рынка жилья и может помочь покупателям, продавцам и специалистам по недвижимости в принятии обоснованных решений. | Price (in rupees) |
| 6 | <https://www.kaggle.com/datasets/sudheerp2147234/salary-dataset-based-on-country-and-race> | Датасет состоит из полной коллекции зарплат и демографической информации с дополнительными сведениями о стаже работы.  Этот датасет представляет собой ценный ресурс для изучения взаимосвязи между доходами и различными социально-демографическими факторами и для проведения комплексных исследований разнообразия доходов и понимания многогранности факторов, влияющих на возможности заработка современной рабочей силы. | Salary |
| 7 | <https://www.kaggle.com/datasets/datasnaek/chess> | Датасет из более 20 000 шахматных партий, собранных от нескольких пользователей на сайте <http://Lichess.org> .  В одной шахматной партии содержится огромное количество информации. Это прежде всего игра с закономерностями, которые наука о данных и изучает. Надо понять, что позволяет игроку выигрывать черными или белыми, насколько сильно влияют на партию мета-факторы (внеигровые), связь между дебютами и победой черных и белых и многое другое.  Создайте модель, которая поможет определить победителя партии. | Winner |
| 8 | <https://www.kaggle.com/datasets/abhijitdahatonde/real-world-smartphones-dataset> | В этом датасете собрана полная информация обо всех новейших смартфонах, доступных на рынке по состоянию на текущий момент времени.  Данный датасет был создан с помощью веб-скраппинга интернет-ресурсов с целью сбора точной и актуальной информации о различных моделях смартфонов, их технических характеристиках, возможностях и ценах.  Найдите закономерности в формировании цены и постройте модель, которая сможет прогнозировать цену. | Price |
| 9 | <https://www.kaggle.com/datasets/kapturovalexander/bank-credit-scoring> | Кредитный скоринг — это система оценки заемщика, с помощью которой банки и могут предсказать, насколько аккуратно человек будет выплачивать кредит.  В данном датасете содержится информация о клиентах. Создайте модель, которая будет предсказывать - выдавать кредит или нет. | y |
| 10 | <https://www.kaggle.com/datasets/juanmerinobermejo/laptops-price-dataset> | В этом датасете собрана полная информация о различных ноутбуках, позволяющая детально проанализировать их технические характеристики и цены. В нем собраны ноутбуки различных марок, моделей и конфигураций.  Создайте модель, которая будет определять конечную цену ноутбука. | Final price |
| 11 | <https://www.kaggle.com/datasets/kellistephenson/increasing-breast-cancer-awareness> | Набор данных содержит информацию о пациентах с раком молочной железы.  Исследование может быть использовано для анализа взаимосвязи между уровнем экспрессии белков, стадией рака и исходом заболевания, понимания влияния различных типов операций на выживаемость пациенток и выявления потенциальных факторов риска прогрессирования рака молочной железы.  Разработайте модель, которая будет определять статус пациента. | Patient status |
| 12 | <https://www.kaggle.com/datasets/karkavelrajaj/amazon-sales-dataset> | Этот набор данных содержит данные о рейтингах и отзывах продуктов Amazon в соответствии с их информацией, указанной на официальном сайте Amazon.  Amazon - это многонациональная технологическая компания, в сферу интересов которой входит электронная коммерция, где она покупает и хранит товарные запасы, а также занимается всем, начиная от доставки и ценообразования и заканчивая обслуживанием клиентов и возвратом товаров.  Разработайте рекомендательную систему на основе данного датасета. | Рекомендательная система. Без признака. |
| 13 | <https://www.kaggle.com/datasets/dansbecker/new-york-city-taxi-fare-prediction> | Датасет содержит информацию о поездках в такси в городе Нью-Йорк.  Ваша задача построить модель для предсказания цены тарифа на такси. | fare\_amount |
| 14 | <https://www.kaggle.com/datasets/hellbuoy/car-price-prediction> | Китайская автомобильная компания Geely Auto стремится выйти на новые рынки, открыв там свое производство и выпуская автомобили на месте, чтобы составить конкуренцию своим американским и европейским коллегам.  Они обратились в автомобильную консалтинговую компанию, чтобы понять, от каких факторов зависит ценообразование на автомобили. В частности, они хотят понять, какие факторы влияют на ценообразование автомобилей на рынке, поскольку они могут сильно отличаться от китайского рынка.  Постройте модель предсказания цены на автомобили. | price |
| 15 | <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link> | Этот датасет содержит информацию о пациентах с раком легких.  Рак легких является ведущей причиной смерти от онкологических заболеваний во всем мире, на долю которой в 2018 году пришлось 1,59 млн смертей. Большинство случаев рака легких связано с курением, однако воздействие загрязнения воздуха также является одним из факторов риска. Новое исследование показало, что загрязнение воздуха может быть связано с повышенным риском развития рака легких даже у некурящих людей.  Разработайте модель, которая будет предсказывать риск развития рака легких. | Level |
| 16 | <https://www.kaggle.com/datasets/uciml/student-alcohol-consumption> | Датасет получен в ходе опроса учащихся математического и португальского языковых курсов в средней школе. В нем содержится информации о студентах.  Создайте модель, которая будет предсказывать итоговую оценку студента. | G3 |
| 17 | <https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets> | Датасет компании Orange Telecom, состоящий из данных об активности клиентов, а также метки churn, указывающей, отменил ли клиент подписку. С помощью данной модели в будущем можно будет создавать специальные программы для удержания клиентов.  Разработайте модель, которая будет предсказывать, отменит ли клиент подписку или нет. | Churn |
| 18 | <https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset> | Датасет содержит информацию о внешности людей.  Постройте модель, которая будет предсказывать какого пола человек, основываясь на его внешности. | Gender |
| 19 | <https://www.kaggle.com/datasets/sgpjesus/bank-account-fraud-dataset-neurips-2022> | На выставке NeurIPS 2022 был опубликован датасет Bank Account Fraud (BAF), состоящий из 6 различных синтетических табличных наборов данных о мошенничестве с банковскими счетами.  Ваша задача - построить модель, которая будет определять, мошенническая ли операция или нет. | fraud\_bool |
|  |  |  |  |
|  |  |  |  |