**9. Классификация типа биома (главного лесного покрова) на основе различных физических характеристик и землепользования.**

*Описание:*

Датасет "Covertype" представляет собой набор данных о типах покрытия леса в горной местности Колорадо, США. Датасет содержит информацию о различных атрибутах, связанных с лесными участками, таких как высота, уклон, аспект, расстояние до ближайших дорог и рек, а также различные статистические показатели, полученные из пиксельных данных электронной карты и спутникового изображения.

*Данные:  
В датасете представлены следующие признаки:*

1. Elevation: высота в метрах над уровнем моря.  
2. Aspect: аспект - направление экспозиции склона лесного участка в градусах северного направления по часовой стрелке от 0 до 360.  
3. Slope: уклон склона лесного участка в градусах.  
4. Horizontal\_Distance\_To\_Hydrology: горизонтальное расстояние от лесного участка до ближайшей воды (реки или озера) в метрах.  
5. Vertical\_Distance\_To\_Hydrology: вертикальное расстояние от лесного участка до ближайшей воды (реки или озера) в метрах.  
6. Horizontal\_Distance\_To\_Roadways: горизонтальное расстояние от лесного участка до ближайшей дороги в метрах.  
7. Hillshade\_9am: интенсивность освещенности склона лесного участка в 9 утра (от 0 до 255).  
8. Hillshade\_Noon: интенсивность освещенности склона лесного участка в полдень (от 0 до 255).  
9. Hillshade\_3pm: интенсивность освещенности склона лесного участка в 3 часа дня (от 0 до 255).  
10. Horizontal\_Distance\_To\_Fire\_Points: горизонтальное расстояние от лесного участка до ближайшего пункта контроля пожара в метрах.  
11. Wilderness\_Area (4 поля): бинарные признаки, указывающие принадлежность лесного участка к одной из четырех областей дикой природы.  
12. Soil\_Type (40 полей): бинарные признаки, указывающие тип почвы на лесном участке.  
13. Cover\_Type: тип покрытия леса, являющийся целевой переменной. Это категориальный признак, принимающий значения от 1 до 7.

1. Загрузка данных: Загрузить датасет Covertype, состоящий из различных признаков, таких как высота над уровнем моря, угол наклона, горизонтальное и вертикальное расстояние к воде и т.д.  
  
2. Изучение данных: Изучить описание датасета, чтобы понять, какие признаки содержатся в данных и какие значения они принимают.  
  
3. Предварительная обработка данных: Провести предварительную обработку данных, включающую удаление выбросов, заполнение пропущенных значений (если есть) и масштабирование признаков.  
  
4. Исследовательский анализ данных: Визуализировать данные с помощью графиков и диаграмм, чтобы лучше понять распределение каждого признака и его связь с целевой переменной (типом биома).  
  
5. Выбор признаков: Определить наиболее значимые признаки с помощью анализа важности признаков, например, используя метод случайного леса.  
  
6. Разделение данных: Разделить данные на обучающую и тестовую выборки

7. Построение модели: Выбрать подходящий алгоритм классификации, такой как случайный лес, логистическая регрессия или метод k-ближайших соседей.  
  
8. Обучение модели: Обучите выбранную модель на обучающих данных, используя настроенные параметры или подбор их с помощью кросс-валидации.  
  
9. Оценка производительности модели: Оценить производительность модели на тестовых данных, используя метрики классификации, такие как точность, полнота и F1-мера.  
  
10. Настройку модели: Попробовать улучшить производительность модели, выполнив настройку ее гиперпараметров, например, с помощью метода перебора сетки или случайного поиска.  
  
11. Оценка важности признаков: Оценить важность признаков в модели, чтобы определить, какие признаки наиболее сильно влияют на предсказания.  
  
12. Повторение: Повторить шаги 7-11 с использованием других алгоритмов классификации, чтобы сравнить их производительность и выбрать наилучшую модель.  
  
13. Оценка переобучения: Оценить уровень переобучения модели, проведя анализ ошибок и разницу между производительностью на обучающей и тестовой выборках.  
  
14. Оценка результатов: Оцените результаты модели, проанализировав полученные значения метрик классификации и интерпретируя их в контексте задачи классификации типов биома.