

ЗАДАНИЕ

В процессе строительства объектов городской среды застройщикам необходимо выполнять условия, заложенные в проектной документации по срокам и стоимости объекта. Так как существуют различные факторы, влияющие на эти 2 показателя, то и задача прогнозирования выполнения задачи в срок и в установленную стоимость является не линейной.

Задание: разработать алгоритм решения задачи прогнозирования сроков завершения строительства и затрат на основе различных исходных данных, таких как: объем проекта, материалы, рабочая сила и погодные условия и пр.

1. Соберите и предварительно обработайте данные, которые включают в себя входные характеристики (объем проекта, материалы, труд, погодные условия и пр.) и выходные данные (дата завершения проекта и стоимость) (максимум 20 баллов за выполнение)
2. Разделите данные на обучающий набор и тестовый набор. Выберите метод машинного обучения, обучите модель на обучающей выборке (максимум 20 баллов за выполнение)
3. Используйте обученную модель, чтобы предсказать дату завершения и стоимость проектов в тестовом наборе (максимум 20 баллов за выполнение)
4. Оцените производительность модели, сравнив прогнозируемые значения с фактическими значениями и вычислив такие показатели, как средняя абсолютная ошибка или R-квадрат (максимум 20 баллов за выполнение)
5. Выберите и используйте модель для прогнозирования даты завершения и затрат для новых проектов путем ввода данных проекта – напишите код для решения этой задачи (максимум 20 баллов за выполнение)

Представленный код должен быть совместим с Python 3.x.

Для задачи прогнозирования сроков завершения строительства и стоимости можно использовать следующие данные:

1. Объем проекта: Описание проекта, размер здания, количество комнат и т. д.
2. Материалы: тип используемых материалов, стоимость материалов, сроки поставки и т. д.
3. Труд: количество рабочих, почасовая ставка, опыт и уровень квалификации рабочих и т. д.
4. Погодные условия: средняя температура, осадки, скорость ветра и т. д.
5. Исторические данные: предыдущие проекты с аналогичным объемом и условиями, даты их завершения и стоимость и т. д.
6. Данные о прогрессе: Текущий прогресс проекта, статус каждого этапа и т. д.

Информация может быть собрана в электронной таблице, где каждая строка представляет отдельный проект, а столбцы представляют собой различные данные, перечисленные выше. Чтобы использовать эти данные с машинным обучением, вы можете разделить их на две части: набор для обучения и набор для тестирования. Учебный набор будет использоваться для обучения модели машинного обучения прогнозированию дат завершения и затрат, а тестовый набор будет использоваться для оценки производительности модели.