

БЛАНК ОТВЕТОВ

страница 3

Отвечать на задачи необходимо полным, развернутым ответом (решением). Пишите аккуратно и разборчиво, соблюдая разметку страницы. Не забудьте указать номер задачи, на которую Вы отвечаете, например, 4.
Условия задачи переписывать не нужно.

(средняя абсолютная ошибка)
MSE - квадратичная ошибка, которая показывает насколько ошибается модель при прогнозировании.
После проверки обучающей модели на тестовом множестве MSE = 0,36. (0-точное совпадение)

3) При прогнозировании с помощью функции predict() по данным
Объем проекта = 7; ставка наем - 320;
прирост = 210; процентная ставка - 21% растущая
средняя стоимость проекта 228,9 дней,
а стоимость 3003 млн. руб.

Данные для своего проекта можно брать через сервис с помощью сервиса. Карту карту, которая позволяет правильно интерпретировать данные. С помощью

Федеральное государственное бюджетное образовательное учреждение высшего образования "НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ"

Открытая многопрофильная олимпиада
«Строительная олимпиада имени Н.С. Стрелецкого»
по профилю Искусственный интеллект в строительстве
(Информационные технологии в строительстве)

ЧИСТОВИК

Вариант № 1-47

(без № варианта работа не проверяется и аннулируется)

При обнаружении в чистовике записей, не относящихся к решаемому варианту, работа не проверяется и аннулируется

БЛАНК ОЦЕНКИ РЕЗУЛЬТАТОВ

Задача №	Служебное поле
1	исходные данные. X/Sx Файл с иск. данными и объяснением (дополнительно задание)
2	програм.ру
3	програм.ру
4	програм.ру
5	програм.ру (программный код)

Итого:

Отвечать на задачи необходимо полным, развернутым ответом (решением). Пишите аккуратно и разборчиво, соблюдая разметку страницы. Не забудьте указать номер задачи, на которую Вы отвечаете, например, 4. Условия задачи переписывать не нужно.

Анализом решения задачи прогнозировали сроки завершения строительства и затрат.

1) Входные данные включают в себя таблицу данных, в которой строки относятся объектным проектам, а столбцы - признакам, на основе которых будет осуществляться прогноз. Целевые данные - сроки и стоимость, они соответствуют каждому проекту.

№ проекта	Объем пр. (габарит)	Материалы (стоимость)	Трудозатраты (чел/ч)	Дата завершения (сут)	Срок (сут)
1	7	305 млн	200 ч	220 сут	3000 млн
2	4	310 млн	202 ч	222 сут	3001 млн
3	7	315 млн	200 ч	226 сут	3002 млн
4	7	320 млн	210 ч	229 сут	3003 млн
5	7	325 млн	214 ч	232 сут	3004 млн

Объем проекта - качественная характеристика. В работе алгоритма необходимо ввести ее числовому формату. Объем проекта оценивается от 0 до 10.

2) Исходные данные делится на обучающий и тестовый набор в соотношении 80/20, где 80% - для обучения, 20% - для тестирования модели. Таким образом, модель будет обучаться на 80% - 4 проекта из 5 и тестироваться на одном. Модель прогнозировали будет построена на основе метода наименьших квадратов регрессии. Метод наименьших квадратов широко применяется для работы с числовыми непрерывными величинами и демонстрирует хорошие показатели корреляции.

Отвечать на задачи необходимо полным, развернутым ответом (решением). Пишите аккуратно и разборчиво, соблюдая разметку страницы. Не забудьте указать номер задачи, на которую Вы отвечаете, например, 4. Условия задачи переписывать не нужно.

Модель линейной регрессии можно получить из библиотеки SciKit learn в Python. Таким образом, после импорта экземпляра класса `model.LinearRegression()`, используется функция `fit` для обучения тренировочной выборки, которая задается с помощью функции `train-test-split` с атрибутом `test_size = 0.2`, что означает размер тестовой выборки в 20%.

3) Использование модели в тестовом наборе происходит с помощью функции `predict()`, где обучение происходит на наборе `X_train`, и на основе полученных прогнозов рассчитывается степень соответствия о результатам с помощью коэффициентов эффективности R^2 и MSE .

4) Производительность модели оценивается с помощью двух показателей: MSE и R^2 . R^2 - коэффициент детерминации, который показывает взаимосвязь и зависимость прогнозируемых данных от исходных факторов. Проверка модели на R^2 на тестовой выборке после обучения на тренировочной. $R^2 = 0.9367$, что, учитывая малое количество базисных данных, является отличным резул-том.

Файл « 1.py»

```
from pyexpat import model
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_squared_error

# Входные данные - признаки в массиве, объем проекта, материалы, труд, погодные условия
X = np.array([[7, 305, 202, 22], [7, 310, 200, 23],
              [7, 315, 206, 22], [7, 320, 210, 24], [7, 325, 214, 25]])
# Выходные данные - срок строительства, стоимость
Y = np.array([[220, 3000], [222, 3001], [226, 3002], [229, 3003], [232, 3004]])
print(Y)
# выделение обучающей и тренировочной выборки в соотношении 80/20
x_test, x_train, y_test, y_train = train_test_split(
    X, Y, test_size=0.7, random_state=1)

print("x_train", x_train)
print("x_test", x_test)
# Создание модели регрессии
model = LinearRegression()
# Обучение на тренировочных данных
model.fit(x_train, y_train)
# Тестирование обученной модели
y_pred = model.predict(x_test)
print("Тестовые данные: ", y_pred)
# Подсчет R2 коэффициента детерминации
r2 = r2_score(y_test, y_pred)
print("R2: ", r2)
# Подсчет коэффициента квадратичной ошибки
mse = mean_squared_error(y_test, y_pred)
print("MSE: ", mse)
# Пользовательские данные
xUserData=np.array([7, 302, 205, 24])
#Расчет результата
xUserResult = model.predict(xUserData)
print(xUserResult)
```

файл «2.xlsx»

ИИС_01_2 - Microsoft Excel

Общий

Входные данные					Выходные данные	
№ Проекта	Объем проекта	Материалы, млн. руб	Труд, человек	Погодные условия, С*	Дата завершения, суток	Стоимость, млн руб
1	7	305	202	22	220	3000
2	7	310	200	23	222	3001
3	7	315	206	22	226	3002
4	7	320	210	24	229	3003
5	7	325	214	25	232	3004
Проверка данных:		Пустые поля	Корректно			
		Текстовые данные	Корректно			
17 Объем проекта - качественная величина (малый, средний, большой), приведенная к числовому формату величина, изменяющаяся от 0 до 10.						
18 Пример	Средний проект	->	7			
19	Большой проект	->	9			
20						
21	Тип данных всех величин - Integer					
22						
23						