



Проект по модулю

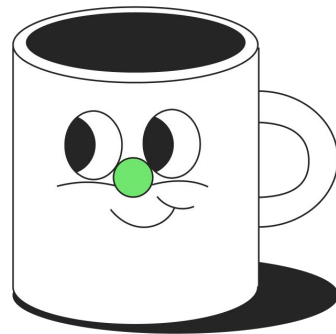
Обзор проекта





Что будет на уроке сегодня

-  Обсудим задание
-  Ответы на вопросы





Задание





Предсказание рака лёгких

Датасет приложен к уроку

- Проведите разведочный анализ данных
- Воспользуйтесь логистической регрессией из пакета `sklearn`
- Оформите результат в виде ноутбука Jupiter

Описание данных

- **GENDER:** M [Male], F [Female]
- **AGE:** Age of patients
- **SMOKING:** 2 [Yes], 1 [No]
- **YELLOW_FINGERS:** 2 [Yes], 1 [No]
- **ANXIETY:** 2 [Yes], 1 [No]
- **PEER_PRESSURE:** 2 [Yes], 1 [No]
- **CHRONIC_DISEASE:** 2 [Yes], 1 [No]
- **FATIGUE:** 2 [Yes], 1 [No]
- **ALLERGY:** 2 [Yes], 1 [No]
- **WHEEZING:** 2 [Yes], 1 [No]
- **ALCOHOL_CONSUMING:** 2 [Yes], 1 [No]
- **COUGHING:** 2 [Yes], 1 [No]
- **SHORTNESS OF BREATH:** 2 [Yes], 1 [No]
- **SWALLOWING DIFFICULTY:** 2 [Yes], 1 [No]
- **CHEST PAIN:** 2 [Yes], 1 [No]
- **LUNG_CANCER:** YES [Positive], NO [Negative]



Как подготовить данные

```
1 from sklearn.preprocessing import LabelEncoder
2 LabelEncoder = LabelEncoder()
3
4 data["GENDER"] = data["GENDER"].replace({"M" : "Male" , "F" : "Female"})
5 data["LUNG_CANCER"] = LabelEncoder.fit_transform(data["LUNG_CANCER"])
6
7 data = pd.get_dummies(data, columns= ["GENDER"])
8 data.rename(columns={"GENDER_Male" : "MALE", "GENDER_Female" : "FEMALE"}, inplace=True)
9
10 data.head()
```



Как выровнять данные

```
1 from sklearn.preprocessing import StandardScaler
2
3 scaler = StandardScaler()
4 x = scaler.fit_transform(x)
```



Как разделить данные

```
1 from sklearn.model_selection import train_test_split
2
3 x_train, x_test, y_train, y_test = train_test_split(x,y,
    test_size=0.2, random_state=0)
```



Как воспользоваться регрессией

```
1 from sklearn.linear_model import LogisticRegression
2 from sklearn.metrics import accuracy_score,
  classification_report
3
4
5 lr = LogisticRegression()
6 lr.fit(x_train, y_train)
7 lr_pred = lr.predict(x_test)
8 lr_report = classification_report(y_test, lr_pred)
9 lr_acc = round(accuracy_score(y_test, lr_pred)*100, ndigits = 2)
10 print(f"\nClassification Report : \n\n{lr_report}")
11 print(f"\nThe Accuracy of Logistic Regression is {lr_acc} %")
```




Вопросы?

Вопросы?



Вопросы?



УДАЧИ