

A background network diagram with white nodes and lines on a dark grey background. The nodes are connected in a complex, interconnected pattern, resembling a social or organizational network. The lines are thin and light grey, while the nodes are small white circles.

CONGRESSIONAL VOTING RECORDS DATASET

Чибисова Ярослава, Петров Артемий 10П

НАЧАЛЬНЫЙ DATAFRAME

	Class Name	handicapped-infants	water-project-cost-sharing	adoption-of-the-budget-resolution	physician-fee-freeze	el-salvador-aid	religious-groups-in-schools	anti-satellite-test-ban	aid-to-nicaraguan-contras	mx-missile	immigration	synfuels-corporation-cutback	education-spending	superfund-right-to-sue	crime	duty-free-exports	export-administration-act-south-africa
0	republican	n	y	n	y	y	y	n	n	n	y	?	y	y	y	n	y
1	republican	n	y	n	y	y	y	n	n	n	n	n	y	y	y	n	?
2	democrat	?	y	y	?	y	y	n	n	n	n	y	n	y	y	n	n
3	democrat	n	y	y	n	?	y	n	n	n	n	y	n	y	n	n	y
4	democrat	y	y	y	n	y	y	n	n	n	n	y	?	y	y	y	y
...
430	republican	n	n	y	y	y	y	n	n	y	y	n	y	y	y	n	y
431	democrat	n	n	y	n	n	n	y	y	y	y	n	n	n	n	n	y
432	republican	n	?	n	y	y	y	n	n	n	n	y	y	y	y	n	y
433	republican	n	n	n	y	y	y	?	?	?	?	n	y	y	y	n	y
434	republican	n	y	n	y	y	y	n	n	n	y	n	y	y	y	?	n

435 rows × 17 columns

Исследование данных

- Исследование данных начинается с определения размерности датафрейма

```
print(df.shape)
## Определяем размерность датафрейма

(435, 17)
```

- После определения размерности, ищем пустые значения(пустых значений в датафрейме нет)

```
[ ] print(df.isna().sum())
## Ищем пустые значения

Class Name      0
handicapped-infants  0
water-project-cost-sharing  0
adoption-of-the-budget-resolution  0
physician-fee-freeze  0
el-salvador-aid  0
religious-groups-in-schools  0
anti-satellite-test-ban  0
aid-to-nicaraguan-contras  0
mx-missile      0
immigration     0
synfuels-corporation-cutback  0
education-spending  0
superfund-right-to-sue  0
crime          0
duty-free-exports  0
export-administration-act-south-africa  0
dtype: int64
```

Обработка данных

- Заменяем категориальные признаки на числовые

```
df['handicapped-infants'] = df['handicapped-infants'].map({'n': 0,
'y': 1})
df['water-project-cost-sharing'] = df['water-project-cost-sharing'].map({'n': 0,
'y': 1})
df['adoption-of-the-budget-resolution'] = df['adoption-of-the-budget-resolution'].map({'n': 0,
'y': 1})
df['physician-fee-freeze'] = df['physician-fee-freeze'].map({'n': 0,
'y': 1})
df['el-salvador-aid'] = df['el-salvador-aid'].map({'n': 0,
'y': 1})
df['religious-groups-in-schools'] = df['religious-groups-in-schools'].map({'n': 0,
'y': 1})
df['anti-satellite-test-ban'] = df['anti-satellite-test-ban'].map({'n': 0,
'y': 1})
df['mx-missile'] = df['mx-missile'].map({'n': 0,
'y': 1})
df['immigration'] = df['immigration'].map({'n': 0,
'y': 1})
df['synfuels-corporation-cutback'] = df['synfuels-corporation-cutback'].map({'n': 0,
'y': 1})
df['education-spending'] = df['education-spending'].map({'n': 0,
'y': 1})
df['superfund-right-to-sue'] = df['superfund-right-to-sue'].map({'n': 0,
'y': 1})
df['crime'] = df['crime'].map({'n': 0,
'y': 1})
df['export-administration-act-south-africa'] = df['export-administration-act-south-africa'].map({'n': 0,
'y': 1})
df['aid-to-nicaraguan-contras'] = df['aid-to-nicaraguan-contras'].map({'n': 0,
'y': 1})
df['duty-free-exports'] = df['duty-free-exports'].map({'n': 0,
'y': 1})
df['Class Name'] = df['Class Name'].map({'republican': 0,
'democrat': 1})
```

Обработка данных

- После замены категориальных признаков в датафрейме появляются пустые

```
df.isna().sum()
# После замены значений появились пустые значения
```

Class Name	0
handicapped-infants	12
water-project-cost-sharing	48
adoption-of-the-budget-resolution	11
physician-fee-freeze	11
el-salvador-aid	15
religious-groups-in-schools	11
anti-satellite-test-ban	14
aid-to-nicaraguan-contras	15
mx-missile	22
immigration	7
synfuels-corporation-cutback	21
education-spending	31
superfund-right-to-sue	25
crime	17
duty-free-exports	28
export-administration-act-south-africa	104

dtype: int64

- Удаляем столбец, в котором слишком много пустых значений, а так же удаляем все столбцы с пустыми значениями

```
[ ] df = df.drop([' export-administration-act-south-africa'], axis=1)
# Удаляем столбец в котором слишком много пустых значений
```

```
[ ] df = df.dropna()
df
# Удаляем все строки с пустыми значениями
```

Обработка данных

- Задаем вектор признаков

```
[ ] X = df.drop(['Class Name'], axis = 1)
    y = df['Class Name']
```

- Масштабируем данные

```
[ ] scaler = MinMaxScaler()
    X_norm = scaler.fit_transform(X)
    X_norm
    # Нормализуем данные
```

```
[ ] scaler = StandardScaler()
    X_st = scaler.fit_transform(X)
    X_st
    # Стандартизируем данные
```