02-main

2023 年 5 月 21 日

1 议员在议案上的投票记录预测其所属党派

为了研究不同党派的议员对议题选择的偏向性，某国国会记录了每个议员对每个议题的投票历史，如图所示。现在想根据某个国会议员对每个议题的投票情况，预测该国会议员属于哪个党派。

1. 完成数据集读取
2. 完成数据集的缺失值填充，把所有放弃投票的值填充为字符串 None
3. 将所有议题作为分类特征
4. 构建分类独热编码器
5. 完成数据分类模型构建及训练
6. 对模型进行 10 折交叉检验
7. 完成数据分类预测

在 Github 中查看

```python
[11]: import pandas as pd
```

```python
[12]: data = pd.read_csv('议案投票.csv')
      data.head()
```

```
[12]:    ID        Name Party Campaign Finance Overhaul
    0   1  Abercrombie     D                          Y  \
    1   2     Ackerman     D                          Y
    2   3     Aderholt     R                          N
    3   4         Akin     R                          N
    4   5     Allen T.     D                          Y


       Unemployment and Tax Benefits Fiscal 2003 Budget Resolution
    0                             N                              N  \
    1                             N                              N
```

```
2                                Y                        Y
3                                Y                        Y
4                                N                        N


   Permanent Tax Cuts Food Stamps Nuclear Waste
0                   N          Y             N  \
1                   N          Y             N
2                   Y          N             Y
3                   Y          N             Y
4                   N          Y             Y


   Fiscal 2003 Defense Authorization  … Estate Tax Repeal
0                                  N  …               Y  \
1                                  N  …               N
2                                  Y  …               Y
3                                  Y  …               Y
4                                  N  …               N


   Married Couples Tax Relief Late Term Abortion Ban Homeland Sec/Union Memb
0                          Y                       N                       N  \
1                          N                       N                       N
2                          Y                       Y                       Y
3                          Y                       Y                       Y
4                          N                       N                       N


   Homeland Sec/Civil Service Emp Homeland Sec/Whistleblower Protections
0                              Y                                       Y  \
1                              Y                                       Y
2                              N                                       N
3                              N                                       N
4                              Y                                       Y


   Andean Trade Abortion Service Refusals Medical Malpractice Awards
0            N                          Y                         N  \
1            N                          Y                         N
2            Y                          N                         Y
```

```
3          Y                    N                    Y
4          N                    Y                    N
```

```
   Military Support for UN Resolution
0                                N
1                                N
2                                N
3                                N
4                                Y
```

```
[5 rows x 23 columns]
```

```
[13]: features = data.columns[3:].tolist()
```

```
[14]: from sklearn.preprocessing import OneHotEncoder
```

```
[15]: one_hot_encoder = OneHotEncoder()
      one_hot_encoder.fit(data[features])
      one_hot_data = one_hot_encoder.transform(data[features])
```

```
[16]: from sklearn.naive_bayes import BernoulliNB
```

```
[17]: BNBModel = BernoulliNB()
```

```
[18]: from sklearn.model_selection import cross_val_score
```

```
[19]: cvs = cross_val_score(
          BNBModel,
          one_hot_data,
          data['Party'],
          cv=10
      )
      cvs.mean()
```

```
[19]: 0.9908562367864693
```