

Noah Caulfield

Dr. Edgar Eduardo Ceh Varela

CS 461

4/11/2024

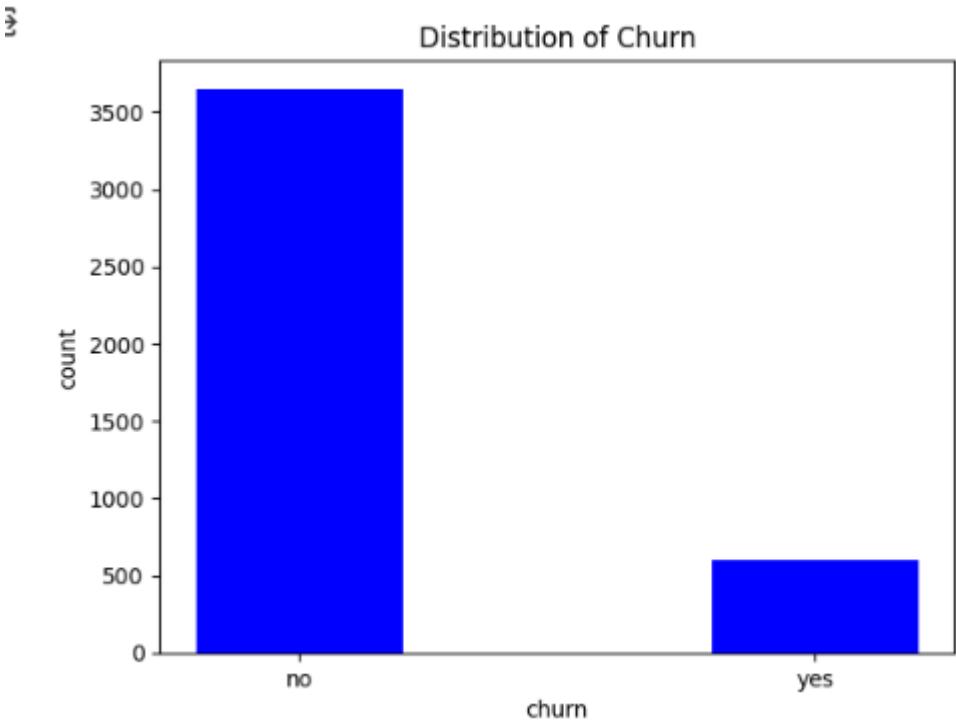
HW5 - Classification using Decision Trees

state account_length area_code international_plan voice_mail_plan number_vmail_messages total_day_minutes total_day_calls total_day_charge total_eve_minutes total_eve_calls total_eve_charge total_night_minut

0	OH	107	area_code_415	no	yes	26	161.6	123	27.47	195.5	103	16.62	25
1	NJ	137	area_code_415	no	no	0	243.4	114	41.38	121.2	110	10.30	16
2	OH	84	area_code_408	yes	no	0	299.4	71	50.90	61.9	88	5.26	19
3	OK	75	area_code_415	yes	no	0	166.7	113	28.34	148.3	122	12.61	18
4	MA	121	area_code_510	no	yes	24	218.2	88	37.09	348.5	108	29.62	21

Next steps: [Generate code with df train](#) [View recommended plots](#)

	account_length	number_vmail_messages	total_day_minutes	total_day_calls	total_day_charge	total_eve_minutes	total_eve_calls	total_eve_charge	total_night_minutes	total_night_calls	total_night_charge	total_intl_l
count	4250.000000	4250.000000	4250.000000	4250.000000	4250.000000	4250.000000	4250.000000	4250.000000	4250.000000	4250.000000	4250.000000	4250
mean	100.236235	7.631765	180.259600	99.907294	30.644682	200.173906	100.176471	17.015012	200.527882	99.839529	9.023892	10
std	39.698401	13.439882	54.012373	19.850817	9.182096	50.249518	19.908591	4.271212	50.353548	20.093220	2.265922	2
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0
25%	73.000000	0.000000	143.325000	87.000000	24.365000	165.925000	87.000000	14.102500	167.225000	86.000000	7.522500	8
50%	100.000000	0.000000	180.450000	100.000000	30.680000	200.700000	100.000000	17.060000	200.450000	100.000000	9.020000	10
75%	127.000000	16.000000	216.200000	113.000000	36.750000	233.775000	114.000000	19.867500	234.700000	113.000000	10.560000	12
max	243.000000	52.000000	351.500000	165.000000	59.760000	359.300000	170.000000	30.540000	395.000000	175.000000	17.770000	20



	state	account_length	area_code	international_plan	voice_mail_plan	number_vmail_messages	total_day_minutes	total_day_calls	total_day_charge	total_eve_minutes	total_eve_calls	total_eve_charge	total_night_minut
0	OH	107	area_code_415	no	yes	26	161.6	123	27.47	195.5	103	16.62	254.4
1	NJ	137	area_code_415	no	no	0	243.4	114	41.38	121.2	110	10.30	162.6
2	OH	84	area_code_408	yes	no	0	299.4	71	50.90	61.9	88	5.26	196.9
3	OK	75	area_code_415	yes	no	0	166.7	113	28.34	148.3	122	12.61	186.9
4	MA	121	area_code_510	no	yes	24	218.2	88	37.09	348.5	108	29.62	212.6

Next steps: [Generate code with X](#) [View recommended plots](#)

	state_AL	state_AR	state_AZ	state_CA	state_CO	state_CT	state_DC	state_DE	state_FL	state_GA	...	state_VA	state_VT	state_WA	state_WI	state_WV	state_WY	area_code_area_code_415	area_code_area_code_510	internati
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0

5 rows x 54 columns

(4250, 54)

	account_length	number_vmail_messages	total_day_minutes	total_day_calls	total_day_charge	total_eve_minutes	total_eve_calls	total_eve_charge	total_night_minutes	total_night_calls	...	state_TN	state_TX	state_UT
0	107	26	161.6	123	27.47	195.5	103	16.62	254.4	103	...	0	0	0
1	137	0	243.4	114	41.38	121.2	110	10.30	162.6	104	...	0	0	0
2	84	0	299.4	71	50.90	61.9	88	5.26	196.9	89	...	0	0	0
3	75	0	166.7	113	28.34	148.3	122	12.61	186.9	121	...	0	0	0
4	121	24	218.2	88	37.09	348.5	108	29.62	212.6	118	...	0	0	0

5 rows x 69 columns

```
Index(['account_length', 'number_vmail_messages', 'total_day_minutes',
      'total_day_calls', 'total_day_charge', 'total_eve_minutes',
      'total_eve_calls', 'total_eve_charge', 'total_night_minutes',
      'total_night_calls', 'total_night_charge', 'total_intl_minutes',
      'total_intl_calls', 'total_intl_charge',
      'number_customer_service_calls'],
      dtype='object')
```

	account_length	number_vmail_messages	total_day_minutes	total_day_calls	total_day_charge	total_eve_minutes	total_eve_calls	total_eve_charge	total_night_minutes	total_night_calls	...	state_VA	state_VT	state_WA
0	107	26	161.6	123	27.47	195.5	103	16.62	254.4	103	...	False	False	False
1	137	0	243.4	114	41.38	121.2	110	10.30	162.6	104	...	False	False	False
2	84	0	299.4	71	50.90	61.9	88	5.26	196.9	89	...	False	False	False
3	75	0	166.7	113	28.34	148.3	122	12.61	186.9	121	...	False	False	False
4	121	24	218.2	88	37.09	348.5	108	29.62	212.6	118	...	False	False	False

5 rows x 69 columns

(3400, 69)
(3400,)
(850, 69)
(850,)

```
DecisionTreeClassifier
DecisionTreeClassifier(random_state=42)
```

```
DecisionTreeClassifier  
DecisionTreeClassifier(random_state=42)
```

Accuracy: 0.9235294117647059

Test score 0.9235

```
DecisionTreeClassifier  
DecisionTreeClassifier(max_depth=5, random_state=42)
```

Accuracy on test data: 0.9235

```
display(dc_classifier)
```

```
DecisionTreeClassifier  
DecisionTreeClassifier(max_depth=10, min_samples_split=30, random_state=42)
```

Accuracy on test data: 0.9329

```
DecisionTreeClassifier  
DecisionTreeClassifier(max_depth=10, min_samples_leaf=2, min_samples_split=30,  
                      random_state=42)
```

Accuracy on test data: 0.9376

Caulfield 4

