home-claim

November 5, 2023

[1]:
```
import pandas as pd
df = pd.read_excel('./Homeowner_Claim_History.xlsx', sheet_name='HOCLAIMDATA')
```

[2]:
```
df.head()
```

[2]:

	policy	exposure	num_claims	amt_claims	f_primary_age_tier	f_primary_gender	f_marital	f_residence_location	f_fire_alarm_type	f_mile_fire_station	f_aoi_tier
0	P00001	1.0	0	0.00	21 - 27	Male	Married	Urban	Alarm Service	< 1 mile	351K - 600K
1	G00002	1.0	0	0.00	38 - 60	Male	Un-Married	Suburban	None	1 - 5 miles	< 100K
2	A00003	1.0	2	3079.01	38 - 60	Female	Married	Suburban	Standalone	< 1 mile	100K - 350K
3	P00004	1.0	1	804.87	28 - 37	Female	Un-Married	Suburban	Standalone	1 - 5 miles	< 100K
4	G00005	1.0	1	638.74	28 - 37	Female	Un-Married	Suburban	Alarm Service	< 1 mile	100K - 350K

[3]:
```
df['Frequency'] = df['num_claims'] / df['exposure']
def group_frequency(frequency):
    if frequency == 0:
        return 0
    elif 0 < frequency <= 1:
        return 1
    elif 1 < frequency <= 2:
        return 2
    elif 2 < frequency <= 3:
        return 3
    else:
        return 4
```

```
df['Frequency Group'] = df['Frequency'].apply(group_frequency)
df = df.dropna(subset=['Frequency Group'])

train_df = df[df['policy'].str.startswith(('A', 'G', 'P'))]
test_df = df[~df['policy'].str.startswith(('A', 'G', 'P'))]
```

[4]:
```
train_df.head()
```

[4]:
```
   policy  exposure  num_claims  amt_claims f_primary_age_tier  \
0  P00001       1.0           0        0.00            21 - 27
1  G00002       1.0           0        0.00            38 - 60
2  A00003       1.0           2     3079.01            38 - 60
3  P00004       1.0           1      804.87            28 - 37
4  G00005       1.0           1      638.74            28 - 37

  f_primary_gender   f_marital f_residence_location f_fire_alarm_type  \
0             Male     Married               Urban     Alarm Service
1             Male  Un-Married            Suburban              None
2           Female     Married            Suburban        Standalone
3           Female  Un-Married            Suburban        Standalone
4           Female  Un-Married            Suburban     Alarm Service

  f_mile_fire_station   f_aoi_tier  Frequency  Frequency Group
0           < 1 mile  351K - 600K        0.0                0
1         1 - 5 miles       < 100K        0.0                0
2           < 1 mile  100K - 350K        2.0                2
3         1 - 5 miles       < 100K        1.0                1
4           < 1 mile  100K - 350K        1.0                1
```

[5]:
```
from itertools import combinations

predictors = ['f_aoi_tier', 'f_fire_alarm_type', 'f_marital',␣
 ↪'f_mile_fire_station', 'f_primary_age_tier', 'f_primary_gender',␣
 ↪'f_residence_location']
all_predictor_combinations = []

for r in range(1, len(predictors) + 1):
    combinations_r = combinations(predictors, r)
    all_predictor_combinations.extend(list(combinations_r))
```

[6]:
```
len(all_predictor_combinations)
```

[6]:
```
127
```

[7]:
```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, mean_squared_error
```

```
import statsmodels.api as sm

aic_list = []
bic_list = []
accuracy_list = []
rmse_list = []

def logisticRegression(predictors):

    X_train = train_df[predictors]
    y_train = train_df['Frequency Group']
    X_train = pd.get_dummies(X_train, columns=predictors, drop_first=True)
    X_train = sm.add_constant(X_train)

    model = sm.MNLogit(y_train, X_train)
    results = model.fit()
    aic_list.append(results.aic)
    bic_list.append(results.bic)

    X_test = test_df[predictors]
    y_test = test_df['Frequency Group']
    X_test = pd.get_dummies(X_test, columns=predictors, drop_first=True)
    X_test = sm.add_constant(X_test)

    y_pred = results.predict(X_test)
    predicted_class = y_pred.idxmax(axis=1)

    accuracy_list.append(accuracy_score(y_test, predicted_class))
    rmse_list.append(mean_squared_error(y_test, predicted_class, squared=False))

for i in range(0,len(all_predictor_combinations)):
    logisticRegression(list(all_predictor_combinations[i]))
```

```
Optimization terminated successfully.
         Current function value: 1.178844
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.177274
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.181948
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.178214
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.168068
         Iterations 7
```

```
Optimization terminated successfully.
         Current function value: 1.181964
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.181629
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.173906
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.178656
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.174881
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.164576
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.178681
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.178332
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.177082
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.173246
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.162929
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.177105
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.176768
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.178030
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.167882
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.181778
         Iterations 6
```

```
Optimization terminated successfully.
         Current function value: 1.181444
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.164134
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.178032
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.177518
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.167897
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.167540
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.181460
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.173710
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.169831
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.159362
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.173744
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.173393
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.174695
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.164387
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.178493
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.178145
         Iterations 6
```

```
Optimization terminated successfully.
         Current function value: 1.160585
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.174705
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.174179
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.164411
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.164036
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.178169
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.173055
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.162736
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.176911
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.176575
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.158892
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.173063
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.172527
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.162760
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.162400
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.176598
         Iterations 6
```

```
Optimization terminated successfully.
         Current function value: 1.163949
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.177846
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.177333
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.167710
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.167355
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.181273
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.163948
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.163445
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.177334
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.167369
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.169637
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.159165
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.173547
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.173197
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.155264
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.169656
         Iterations 6
```

```
Optimization terminated successfully.
         Current function value: 1.169106
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.159199
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.158821
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.173230
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.160398
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.174519
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.173992
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.164222
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.163849
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.177981
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.160407
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.159890
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.174003
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.163872
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.158699
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.172871
         Iterations 6
```

```
Optimization terminated successfully.
         Current function value: 1.172334
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.162565
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.162207
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.176405
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.158708
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.158180
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.172343
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.162231
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.163762
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.163259
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.177148
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.167183
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.163256
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.155068
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.169461
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.168910
         Iterations 6
```

```
Optimization terminated successfully.
         Current function value: 1.159001
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.158624
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.173033
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.155088
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.154545
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.168929
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.158659
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.160218
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.159703
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.173815
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.163683
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.159709
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.158514
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.157986
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.172149
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.162036
         Iterations 7
```

```
Optimization terminated successfully.
         Current function value: 1.157993
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.163069
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.154891
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.154348
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.168733
         Iterations 6
Optimization terminated successfully.
         Current function value: 1.158461
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.154366
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.159520
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.157798
         Iterations 7
Optimization terminated successfully.
         Current function value: 1.154168
         Iterations 7
```

1 a

[8]:
```
train_group_counts = train_df['Frequency Group'].value_counts()
```

[9]:
```
train_group_counts
```

[9]:
```
0    11336
1     5286
2     2423
4      927
3      689
Name: Frequency Group, dtype: int64
```

[10]:
```
test_group_counts = test_df['Frequency Group'].value_counts()
```

[11]:
```
test_group_counts
```

[11]:
```
0    3858
1    1750
2     779
4     251
3     214
Name: Frequency Group, dtype: int64
```

2 b

[12]:
```
min_aic = min(aic_list)
min_aic_index = aic_list.index(min_aic)
print('The lowest AIC value on the Training partition is: ',min_aic,'\nThe␣
  ↪model with the predictors ',all_predictor_combinations[min_aic_index],' is␣
  ↪producing this value')
```

```
The lowest AIC value on the Training partition is:  47836.117643362864
The model with the predictors  ('f_aoi_tier', 'f_fire_alarm_type',
'f_mile_fire_station', 'f_primary_age_tier', 'f_residence_location')  is
producing this value
```

3 c

[13]:
```
min_bic = min(bic_list)
min_bic_index = bic_list.index(min_bic)
print('The lowest BIC value on the Training partition is: ',min_bic,'\nThe␣
  ↪model with the predictors ',all_predictor_combinations[min_bic_index],' is␣
  ↪producing this value')
```

```
The lowest BIC value on the Training partition is:  48285.16616641166
The model with the predictors  ('f_fire_alarm_type', 'f_mile_fire_station',
'f_primary_age_tier')  is producing this value
```

4 d

[14]:
```
max_accuracy = max(accuracy_list)
max_accuracy_index = accuracy_list.index(max_accuracy)
print('The highest Accuracy value on the Testing partition is:␣
  ↪',max_accuracy,'\nThe model with the predictors␣
  ↪',all_predictor_combinations[max_accuracy_index],' is producing this value')
```

```
The highest Accuracy value on the Testing partition is:  0.5640688849970812
The model with the predictors  ('f_aoi_tier', 'f_fire_alarm_type', 'f_marital',
'f_mile_fire_station', 'f_primary_age_tier')  is producing this value
```

5 e

[15]:
```
min_rmse = min(rmse_list)
min_rmse_index = rmse_list.index(min_rmse)
print('he lowest Root Average Squared Error value on the Testing partition is:␣
  ↪',min_rmse,'\nThe model with the predictors␣
  ↪',all_predictor_combinations[min_bic_index],' is producing this value')
```

```
he lowest Root Average Squared Error value on the Testing partition is:
1.2514439995713142
The model with the predictors  ('f_fire_alarm_type', 'f_mile_fire_station',
'f_primary_age_tier')  is producing this value
```

[ ]:
```

```