**Bike\_Sharing\_Demand**

자전거 공유 시스템은 회원 가입, 대여 및 자전거 반납 과정이 도시 전역의 키오스크 위치 네트워크를 통해 자동화되는 자전거 대여 수단입니다. 이러한 시스템을 사용하여 사람들은 한 위치에서 자전거를 대여하고 필요에 따라 다른 장소로 반납할 수 있습니다. 현재 전 세계적으로 500개 이상의 자전거 공유 프로그램이 있습니다.

이러한 시스템에서 생성된 데이터는 이동 기간, 출발 위치, 도착 위치 및 경과 시간이 명시적으로 기록되기 때문에 연구자에게 매력적입니다. 따라서 자전거 공유 시스템은 도시의 이동성을 연구하는 데 사용할 수 있는 센서 네트워크로 기능합니다. 이 대회에서 참가자들은 워싱턴 D.C.의 Capital Bikeshare 프로그램에서 자전거 대여 수요를 예측하기 위해 과거 사용 패턴과 날씨 데이터를 결합하도록 요청받습니다.

|  |  |  |  |
| --- | --- | --- | --- |
| **케이스 처리 요약** | | | |
|  | | N | 퍼센트 |
| 보기 | 학습 | 7610 | 69.9% |
| 검증 | 3276 | 30.1% |
| 유효 | | 10886 | 100.0% |
| 제외됨 | | 0 |  |
| 전체 | | 10886 |  |
|  |  |  |  |
| 모델 요약a,b | | |  |
| 알파 | 학습 R 제곱 | 홀드아웃 R 제곱 |  |
| 1.000 | 0.650 | 0.624 |  |
| a. 종속변수: count | | |  |
| b. 모델: temp, atemp, humidity, windspeed, hour, season, holiday, workingday, weather, weekday | | |  |

라쏘 회귀분석을 진행하였습니다. 전체 데이터는 10886개 이며 알파가 1일 때 모델의 유의성은 65%였으며, 라소 회귀 검정 결과 30.1% 의 성능을 보였습니다.

|  |  |  |  |
| --- | --- | --- | --- |
| **케이스 처리 요약** | | | |
|  | | N | 퍼센트 |
| 보기 | 학습 | 7655 | 70.3% |
| 검증 | 3231 | 29.7% |
| 유효 | | 10886 | 100.0% |
| 제외됨 | | 0 |  |
| 전체 | | 10886 |  |
|  |  |  |  |
| 모델 요약a,b | | |  |
| 알파 | 학습 R 제곱 | 홀드아웃 R 제곱 |  |
| 1.000 | 0.747 | 0.586 |  |
| a. 종속변수: count | | |  |
| b. 모델: temp, atemp, humidity, windspeed, hour, season, holiday, workingday, weather, weekday | | |  |

리지 회귀분석을 진행하였습니다. 알파가 1일 때 모델의 유의성은 74.7% 였으며, 29.7%의 성능을 보였습니다.

|  |  |  |  |
| --- | --- | --- | --- |
| **케이스 처리 요약** | | | |
|  | | N | 퍼센트 |
| 보기 | 학습 | 7624 | 70.0% |
| 검증 | 3262 | 30.0% |
| 유효 | | 10886 | 100.0% |
| 제외됨 | | 0 |  |
| 전체 | | 10886 |  |
|  |  |  |  |
| 모델 요약a,b | | |  |
| L1 비율 | 알파 | 학습 R 제곱 |  |
| .500 | 1.000 | 0.602 |  |
| a. 종속변수: count | | |  |
| b. 모델: temp, atemp, humidity, windspeed, hour, season, holiday, workingday, weather, weekday | | |  |

엘라스틱 넷 회귀분석을 진행하였습니다. 모델의 유의성은 60.2% 였으며, 성능은 30.0% 를 보였습니다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 라소 회귀 계수a | | | | | |
| 알파 |  | 평균 | 표준 편차 | 표준화 계수 | 비표준화 계수 |
| 1.000 | 절편b |  |  | 189.902 | 183.192 |
|  | [hour=17] | 0.042 | 0.202 | 45.170 | 224.056 |
|  | [hour=18] | 0.041 | 0.199 | 38.064 | 191.375 |
|  | [hour=8] | 0.041 | 0.198 | 31.996 | 161.363 |
|  | season | 2.506 | 1.117 | 21.639 | 19.379 |
|  | [atemp=31.060] | 0.062 | 0.240 | 20.054 | 83.392 |
|  | [hour=19] | 0.043 | 0.202 | 17.648 | 87.409 |
|  | [hour=16] | 0.042 | 0.200 | 15.219 | 75.940 |
|  | [temp=31.16] | 0.023 | 0.151 | 13.005 | 86.285 |
|  | [temp=29.52] | 0.033 | 0.178 | 10.825 | 60.968 |
|  | [temp=30.34] | 0.027 | 0.162 | 8.901 | 55.107 |
|  | [atemp=32.575] | 0.025 | 0.157 | 8.599 | 54.832 |
|  | [temp=32.80] | 0.017 | 0.128 | 8.590 | 67.056 |
|  | [atemp=31.820] | 0.027 | 0.162 | 8.433 | 52.090 |
|  | [temp=33.62] | 0.013 | 0.113 | 6.741 | 59.786 |
|  | [temp=28.70] | 0.041 | 0.198 | 6.608 | 33.323 |
|  | [atemp=30.305] | 0.032 | 0.176 | 6.289 | 35.700 |
|  | [temp=22.96] | 0.038 | 0.192 | 5.631 | 29.365 |
|  | [temp=31.98] | 0.008 | 0.090 | 5.013 | 55.763 |
|  | [temp=23.78] | 0.018 | 0.134 | 4.822 | 36.010 |
|  | [hour=13] | 0.041 | 0.199 | 4.433 | 22.289 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 리지 회귀분석 계수a | | | | | |
| 알파 |  | 평균 | 표준 편차 | 표준화 계수 | 비표준화 계수 |
| 1.000 | 절편b |  |  | 93.230 | 63.943 |
|  | [hour=17] | 0.046 | 0.210 | 25.738 | 122.340 |
|  | [hour=8] | 0.039 | 0.195 | 21.816 | 112.166 |
|  | [hour=18] | 0.038 | 0.191 | 21.716 | 113.615 |
|  | season | 1.342 | 0.475 | 12.550 | 26.447 |
|  | [hour=19] | 0.044 | 0.206 | 10.837 | 52.664 |
|  | [atemp=31.060] | 0.038 | 0.191 | 9.853 | 51.548 |
|  | [temp=24.60] | 0.018 | 0.134 | 9.689 | 72.316 |
|  | [humidity=38] | 0.013 | 0.112 | 7.709 | 68.958 |
|  | [atemp=33.335] | 0.004 | 0.059 | 7.323 | 123.720 |
|  | [hour=9] | 0.043 | 0.203 | 7.085 | 34.965 |
|  | [temp=28.70] | 0.006 | 0.075 | 7.064 | 94.444 |
|  | workingday | 0.695 | 0.460 | 6.332 | 13.750 |
|  | [hour=16] | 0.043 | 0.203 | 6.239 | 30.790 |
|  | [temp=22.14] | 0.025 | 0.155 | 5.775 | 37.273 |
|  | [atemp=25.760] | 0.025 | 0.155 | 5.775 | 37.273 |
|  | [atemp=32.575] | 0.004 | 0.065 | 5.387 | 83.108 |
|  | [humidity=33] | 0.013 | 0.112 | 5.200 | 46.513 |
|  | [temp=26.24] | 0.008 | 0.091 | 5.155 | 56.356 |
|  | [atemp=24.240] | 0.025 | 0.155 | 5.091 | 32.857 |
|  | [temp=20.50] | 0.025 | 0.155 | 5.091 | 32.857 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 엘라스틱 넷 회귀 계수a | | | | | | |
| L1 비율 | 알파 |  | 평균 | 표준 편차 | 표준화 계수 | 비표준화 계수 |
| .500 | 1.000 | 절편b |  |  | 170.351 | 180.756 |
|  |  | [hour=17] | 0.041 | 0.199 | 29.197 | 146.715 |
|  |  | [hour=18] | 0.041 | 0.198 | 24.789 | 125.310 |
|  |  | [hour=8] | 0.042 | 0.201 | 23.532 | 116.871 |
|  |  | [atemp=31.060] | 0.069 | 0.254 | 12.873 | 50.745 |
|  |  | [hour=19] | 0.042 | 0.201 | 12.583 | 62.735 |
|  |  | [hour=16] | 0.042 | 0.200 | 9.981 | 49.860 |
|  |  | [atemp=32.575] | 0.025 | 0.155 | 7.438 | 48.035 |
|  |  | [atemp=31.820] | 0.026 | 0.158 | 7.036 | 44.401 |
|  |  | [temp=32.80] | 0.015 | 0.122 | 6.010 | 49.131 |
|  |  | [temp=29.52] | 0.026 | 0.160 | 5.913 | 36.842 |
|  |  | [hour=9] | 0.045 | 0.208 | 5.505 | 26.442 |
|  |  | [temp=31.16] | 0.020 | 0.139 | 5.344 | 38.522 |
|  |  | [hour=12] | 0.041 | 0.198 | 5.203 | 26.302 |
|  |  | [temp=28.70] | 0.036 | 0.187 | 5.089 | 27.248 |
|  |  | [hour=13] | 0.044 | 0.204 | 4.856 | 23.755 |
|  |  | [temp=24.60] | 0.037 | 0.190 | 4.529 | 23.873 |
|  |  | [temp=34.44] | 0.008 | 0.086 | 4.482 | 51.916 |
|  |  | [temp=26.24] | 0.041 | 0.199 | 4.397 | 22.092 |
|  |  | [humidity=30] | 0.006 | 0.078 | 4.262 | 54.534 |
|  |  | [temp=23.78] | 0.021 | 0.142 | 4.154 | 29.206 |

위에는 각 라쏘 회귀, 리지 회귀, 엘라스틱 넷 회귀의 회귀 계수들을 표준화 계수(B)의 20개 상위값을 표시한 표입니다.

3가지 모델에서 상위 30개에 모두 포함된 변수들은 다음과 같습니다.

|  |
| --- |
| 중요변수 |
| [hour=17] |
| [hour=18] |
| [hour=8] |
| [atemp=31.060] |
| [hour=19] |
| [hour=16] |
| [atemp=32.575] |
| [temp=28.70] |

위에 순서대로 해당하는 변수들이 공유 자전거 수요에 영향을 미쳤음을 알 수 있었습니다.

**추가 개선 사항**

라쏘와 리지, 엘라스틱 넷 회귀 모두 일반회귀모형의 다중공선성을 방지하거나 범주형 변수와 연속형 변수가 동시에 있을 때 적용하는 방법의 회귀 모형입니다.

학습데이터의 경우 모두 255개의 변수들로 이루어진 모형으로 만들어졌습니다. 모형에 비해 너무 변수가 많은 것을 확인하였습니다.

변수를 제거하기보다는 변수를 범주화 시켜서 몇가지의 경우로 줄이는 것을 권장합니다.

조금 더 정확한 분석을 위해서는 학습데이터의 연속형 변수들을 범주화하여 다시 모델을 학습시켜야 할 것입니다.