

YAPAY SİNİR AĞLARI İLE HAVA DURUMU/SICAKLIK TAHMİNLEME

ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ FEN EDEBİYAT FAKÜLTESİ MATEMATİK BİLGİSAYAR BÖLÜMÜ
YAPAY AĞLAR DERSİ 2019-2020 BAHAR YARIYILI
YILSONU PROJESİ

HAZIRLAYAN
MUSTAFA TOSUN, 121620123067
Mayıs 2020

ÖZET: Bu çalışmada, Waikato Environment for Knowledge Analysis (Weka) yazılımı ve yazılımdaki M5P ağaç (*tree*) metodu kullanılarak Yapay Sinir Ağları (YSA) ile hava sıcaklığı tahminlemesi yapılmıştır. Hava sıcaklığının en düşük hata ile tespit edilmesi amaçlanmıştır. Gerçek meteorolojik veriler, .NET 4.5 ortamında C# programlama dili ile toplanmış ve Weka yazılımına aktarılacak üzere işlenmiştir.

1. Projedeki Problem

Belirli bir lke, blge veya merkezde, bir zaman dilimi iinde grlebilecek meteorolojik olayların gzlem ve analizlere dayanılarak sbjektif veya objektif yntemler kullanılarak nceden ngrlme alıřmaları hava tahmini olarak adlandırılır. Hava tahmini  ařamalıdır: Gzlemler, Analiz ve Tahmin.¹

Bu projede, hava sıcaklıęı tahmini iin sbjektif yntemlere ihtiya duyuluyor olması durumundan yola ıkılarak YSA kullanımıyla bu deęerlerin, insan deęiřkenini ortadan kaldırarak tahmin ettirilmesi amalanmıřtır.

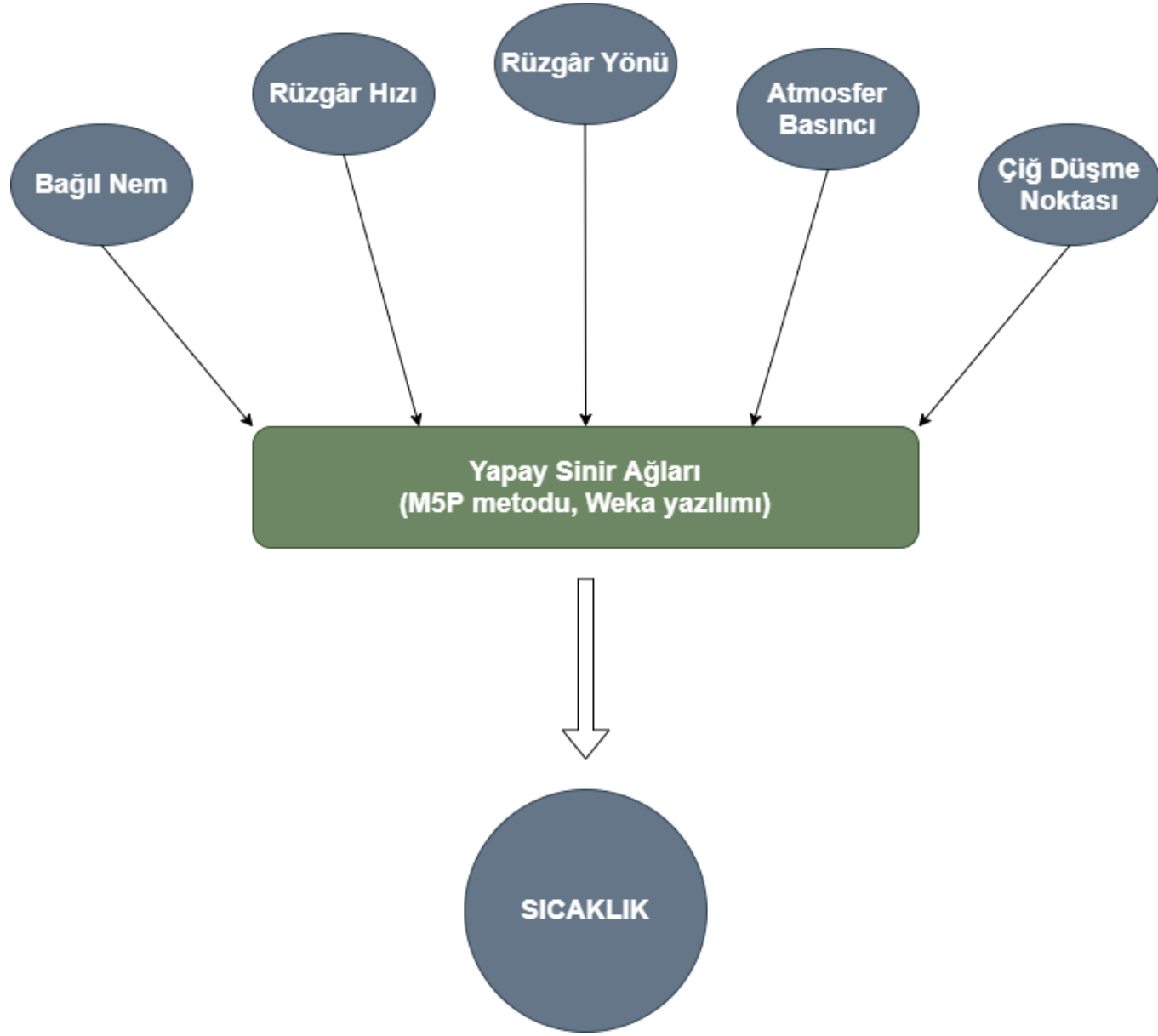
2. zm nerisi

Hava sıcaklıęı hesaplamasında kullanılan objektif argmanların řunlar olduęu tespit edilmiřtir:

- Baęıl nem,
- Rzgr hızı ve yn,
- Atmosfer basıncı ve
- Su buharı basıncı ya da ię dřme noktası

Su buharı basıncı ile ię dřme noktası (iřb sıcaklıęı) arasında doęrudan iliřki olmasından dolayı, bu projede su buharı basıncı yerine ię dřme noktası argmanının kullanımı tercih edilmiřtir.

Bylelikle sistemin giriř ve ıkıř parametrelerini gsteren diyagram řyledir:



Verilerin toplanması aşamasında <https://tr.freemeteo.com/> internet sitesinden yararlanılmıştır. Eskişehir şehir merkezi için son 12 yıla ait günlük meteorolojik veriler toplanmış ve Weka yazılımına gönderilmek üzere filtrelendiği. Veri toplama işlemi sırasında bazı günlere ait hava durumu verilerinin kısmen veya tamamen eksik olduğu tespit edilmiş ve bu eksik veriler veri setine dâhil edilmemiştir.

Günlük hava durumu verilerini filtreleme aşamasında sırasıyla şu işlemler yapılmıştır:

- Rüzgâr hızı tespit edilememiş olan veriler, veri setinden çıkarılmıştır.
- Rüzgâr yönü bulunmayan verilerin rüzgâr yönünün "değişken" olduğu tespit edilmiş ve veri setine bu şekilde uyarlanarak dâhil edilmiştir.
- Verilerin tahmin saatleri 15'er dakikalık dilimlere yuvarlanmıştır.
- Veriler, yuvarlanmış tahmin saatlerine göre gruplandırılmıştır.
- Veriler, hava durumu veri toplayıcısı istasyonlarına göre gruplandırılmıştır.
- Yukarıdaki diyagramda görsellenmiş olan sistemdeki giriş ve çıkış parametreleri birebir aynı olan verilerin yalnızca bir tanesi seçilmiştir.
- En fazla veriye sahip saat grubu ve istasyon grubu verileri bu proje için temel alınmak üzere seçilmiştir.

Bu filtreleme sayesinde; hava sıcaklığına etki edebilecek olan coğrafi konum/desen ve Güneş'in bağıl konumu gibi etmenlerin ortadan kaldırılması amaçlanmıştır. Mevsimsel değişkenler göz ardı edilmiştir.

Filtrelenen veriler Weka yazılımına, eğitim veri seti ve test veri seti olarak iki grupta aktararak, hava sıcaklığı tahmini yaptırılmıştır.

3. Veri Seti Bilgisi

Sistemin bir giriş parametresi olan Rüzgâr Yönü argümanı, 10 derecelik aralıklarla tamsayı olarak (minimum 10, maksimum 360) elde edilmiş olup bunlara ek olarak rüzgâr yönünün “değişken (*variable*)” olabileceği tespit edilmiştir. Dolayısıyla, sistemdeki diğer tüm parametreler Weka yazılımı için “numeric” değerler olarak değerlendirilirken, Rüzgâr Yönü parametresi {Variable, 10, 20, 30, ..., 360} değerlerini içeren “nominal” bir değer olarak alınmıştır.

Filtreleme işleminden sonra yuvarlanmış saat grubu 15:45 ve hava durumu istasyonu Eskişehir isimli istasyon (*Enlem 39.783, Boylam 30.567, Yükseklik 786m*) olan, 25 Temmuz 2008 ve 28 Mayıs 2020 kapalı aralığında bulunan 1180 güne ait (dolayısıyla 1180 adet veri) birbirinden farklı verilerin bulunduğu bir veri seti elde edilmiştir.

Filtrelenen verilerin, Weka yazılımına gönderilmek üzere; %10'luk kısmı (118 adet) test verisi olarak rastgele seçilmiş, geriye kalan %90'luk kısım (1062 adet) ise eğitim verisi olarak iki gruba ayrılmıştır.

Elde edilen eğitim ve test veri setlerinin bir kısım içeriği şöyle listelenmiştir:

| EĞİTİM VERİLERİ | | | | | | | TEST VERİLERİ | | | | | | |
|-----------------|---------------|------------------|-------------------------|----------------------|-----------------|--------------------|---------------|---------------|------------------|-------------------------|----------------------|-----------------|--------------------|
| | A | B | C | D | E | F | | A | B | C | D | E | F |
| 1 | WindSpeedKmph | WindDirection360 | RelativeHumidityPercent | AmosphericPressureMb | DewPointCelcius | TemperatureCelcius | 1 | WindSpeedKmph | WindDirection360 | RelativeHumidityPercent | AmosphericPressureMb | DewPointCelcius | TemperatureCelcius |
| 2 | 19 | 300 | 55 | 1015 | 6 | 15 | 2 | 19 | 20 | 51 | 1018 | 4 | 14 |
| 3 | 24 | 290 | 59 | 1013 | 8 | 16 | 3 | 24 | 200 | 42 | 1002 | 4 | 17 |
| 4 | 6 | 270 | 77 | 1018 | 8 | 12 | 4 | 33 | 300 | 40 | 1010 | 13 | 28 |
| 5 | 19 | 290 | 49 | 1019 | 7 | 18 | 5 | 7 | 320 | 33 | 1012 | 12 | 30 |
| 6 | 15 | 300 | 34 | 1021 | 2 | 18 | 6 | 11 | 270 | 33 | 1016 | 4 | 21 |
| 7 | 22 | 310 | 54 | 1021 | 2 | 11 | 7 | 17 | 100 | 49 | 1021 | 7 | 18 |
| 8 | 33 | 220 | 18 | 1005 | 5 | 32 | 8 | 11 | 310 | 34 | 1019 | 7 | 24 |
| 9 | 11 | 320 | 22 | 1019 | 8 | 32 | 9 | 11 | 320 | 66 | 1015 | 3 | 9 |
| 10 | 17 | 290 | 25 | 1021 | 9 | 31 | 10 | 6 | 80 | 35 | 1025 | 3 | 19 |
| 11 | 22 | 300 | 24 | 1021 | 8 | 31 | 11 | 19 | 130 | 76 | 999 | 4 | 8 |
| 12 | 9 | Variable | 19 | 1019 | 7 | 34 | 12 | 19 | 300 | 17 | 1015 | 2 | 30 |
| 13 | 15 | | 310 | 1019 | 7 | 30 | 13 | 24 | 310 | 71 | 1017 | 4 | 9 |
| 14 | 15 | 250 | 22 | 1017 | 2 | 25 | 14 | 11 | 100 | 71 | 1024 | 1 | 6 |
| 15 | 11 | 330 | 20 | 1017 | -2 | 22 | 15 | 13 | Variable | 31 | 1015 | 10 | 29 |
| 16 | 11 | 330 | 39 | 1017 | 2 | 16 | 16 | 6 | | 270 | 93 | 1021 | 3 |
| 17 | 13 | 260 | 87 | 1014 | 7 | 9 | 17 | 20 | 330 | 75 | 1011 | -3 | 1 |
| 18 | 15 | 250 | 82 | 1014 | 7 | 10 | 18 | 7 | 100 | 54 | 1018 | 1 | 10 |
| 19 | 9 | 120 | 88 | 1008 | 9 | 11 | 19 | 2 | Variable | 15 | 1014 | 3 | 33 |
| 20 | 17 | 130 | 63 | 1005 | 8 | 15 | 20 | 9 | | 280 | 66 | 1019 | 0 |
| 21 | 17 | 320 | 49 | 1009 | 6 | 17 | 21 | 17 | 250 | 35 | 1017 | 5 | 21 |
| 22 | 13 | 260 | 49 | 1011 | 6 | 17 | 22 | 4 | Variable | 36 | 1016 | 9 | 25 |

Veri setleri, Weka yazılımına gönderilmek üzere .*arff* dosya formatında kaydedilmiştir.

4. Eğitim Süreci ve Algoritma

Yapay sinir ağının eğitiminde ve test aşamasında kullanılmak üzere sağlanan veri setleri, 5 girişli (*input*) 1 çıkışlı (*output*) yapay sinir ağında kullanılmıştır. Verilen eğitim veri seti ile yapay sinir ağının öğrenmesi sağlandıktan sonra başarı düzeyini test etmek ve sıcaklık tahmin kabiliyetini ölçmek için test veri seti yapay sinir ağına verilmiştir.

Sunulan girdiye bağlı olarak ağın çıktı değerleri hesaplanır. Eğitim verilerinden yola çıkarak çıkış verileri tahmin edilmeye çalışılır. Hesaplama en temel amaç en düşük hata ile çıkış değerlerinin tahmin edilmeye çalışılmasıdır.

Weka, çok sayıda regresyon ve sınıflandırma araçları içermektedir. Yazılımın native (yerel/doğal/birinci parti) paketleri arasında en ünlü aracı (tool) **M5P model tree** paketidir. Ross J. Quinlan tarafından geliştirilen M5 algoritmasına iyileştirmeler yapılarak Yong Wang tarafından elde edilmiş olan M5P algoritması, etki alanını ardışık ikili (binary) bölgelere ayırır ve sonra her bir ağaç düğümüne (node) lineer modeller yerleştirir.

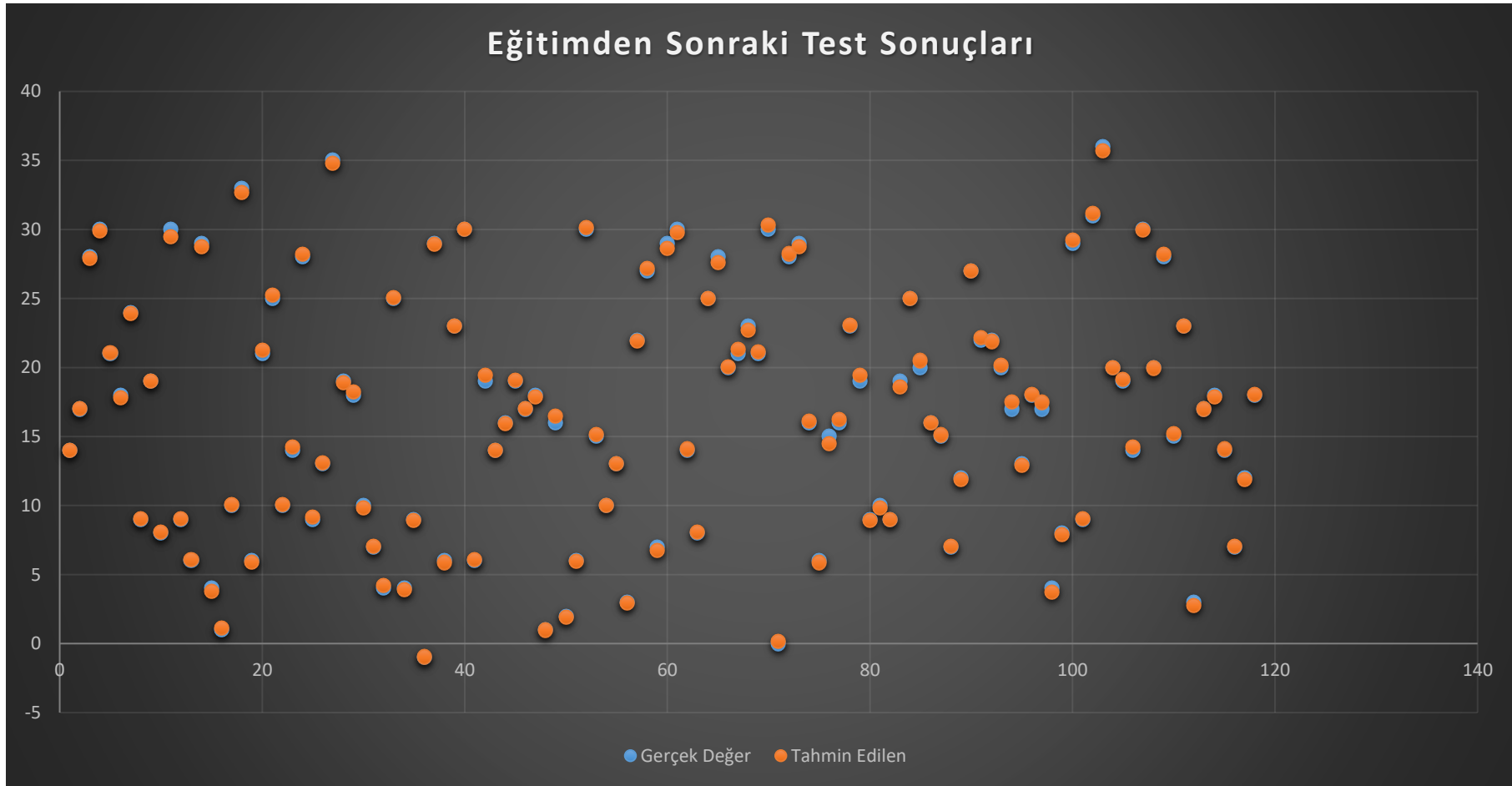
Bu projede, Weka yazılımı ile kullanılan YSA modeli için **M5P** ağaç (*tree*) classifier'i tercih edilmiştir. Fonksiyon parametreleri şöyledir:

```
M5P -M 4.0 -num-decimal-places 4
```

5. Sonular

Yapay sinir ađının eđitiminden sonra verilen test veri seti iin yapılan tahminleme sonucunda;

- Ortalama Mutlak Hata : **0,1484** ve
- Bađıl Mutlak Hata : **%2,0277** olarak elde edilmiřtir.



| # | Gerçek Değer | Tahmin Edilen | Hata |
|----|--------------|---------------|--------|
| 1 | 14 | 14,008 | 0,008 |
| 2 | 17 | 17,025 | 0,025 |
| 3 | 28 | 27,915 | -0,085 |
| 4 | 30 | 29,886 | -0,114 |
| 5 | 21 | 21,078 | 0,078 |
| 6 | 18 | 17,829 | -0,171 |
| 7 | 24 | 23,914 | -0,086 |
| 8 | 9 | 9,039 | 0,039 |
| 9 | 19 | 19,008 | 0,008 |
| 10 | 8 | 8,071 | 0,071 |
| 11 | 30 | 29,495 | -0,505 |
| 12 | 9 | 9,027 | 0,027 |
| 13 | 6 | 6,086 | 0,086 |
| 14 | 29 | 28,758 | -0,242 |
| 15 | 4 | 3,758 | -0,242 |
| 16 | 1 | 1,099 | 0,099 |
| 17 | 10 | 10,044 | 0,044 |
| 18 | 33 | 32,706 | -0,294 |
| 19 | 6 | 5,934 | -0,066 |
| 20 | 21 | 21,277 | 0,277 |
| 21 | 25 | 25,26 | 0,26 |
| 22 | 10 | 10,089 | 0,089 |
| 23 | 14 | 14,229 | 0,229 |
| 24 | 28 | 28,205 | 0,205 |
| 25 | 9 | 9,14 | 0,14 |
| 26 | 13 | 13,112 | 0,112 |
| 27 | 35 | 34,772 | -0,228 |
| 28 | 19 | 18,865 | -0,135 |
| 29 | 18 | 18,221 | 0,221 |
| 30 | 10 | 9,843 | -0,157 |

| # | Gerçek Değer | Tahmin Edilen | Hata |
|----|--------------|---------------|--------|
| 31 | 7 | 7,062 | 0,062 |
| 32 | 4 | 4,223 | 0,223 |
| 33 | 25 | 25,035 | 0,035 |
| 34 | 4 | 3,904 | -0,096 |
| 35 | 9 | 8,93 | -0,07 |
| 36 | -1 | -0,96 | 0,04 |
| 37 | 29 | 28,928 | -0,072 |
| 38 | 6 | 5,876 | -0,124 |
| 39 | 23 | 22,993 | -0,007 |
| 40 | 30 | 29,989 | -0,011 |
| 41 | 6 | 6,076 | 0,076 |
| 42 | 19 | 19,415 | 0,415 |
| 43 | 14 | 13,966 | -0,034 |
| 44 | 16 | 15,955 | -0,045 |
| 45 | 19 | 19,076 | 0,076 |
| 46 | 17 | 17,04 | 0,04 |
| 47 | 18 | 17,882 | -0,118 |
| 48 | 1 | 0,969 | -0,031 |
| 49 | 16 | 16,478 | 0,478 |
| 50 | 2 | 1,899 | -0,101 |
| 51 | 6 | 5,949 | -0,051 |
| 52 | 30 | 30,154 | 0,154 |
| 53 | 15 | 15,133 | 0,133 |
| 54 | 10 | 10,031 | 0,031 |
| 55 | 13 | 13,009 | 0,009 |
| 56 | 3 | 2,95 | -0,05 |
| 57 | 22 | 21,914 | -0,086 |
| 58 | 27 | 27,144 | 0,144 |
| 59 | 7 | 6,75 | -0,25 |
| 60 | 29 | 28,599 | -0,401 |

| # | Gerçek Değer | Tahmin Edilen | Hata |
|----|--------------|---------------|--------|
| 61 | 30 | 29,799 | -0,201 |
| 62 | 14 | 14,085 | 0,085 |
| 63 | 8 | 8,05 | 0,05 |
| 64 | 25 | 24,997 | -0,003 |
| 65 | 28 | 27,582 | -0,418 |
| 66 | 20 | 20,03 | 0,03 |
| 67 | 21 | 21,299 | 0,299 |
| 68 | 23 | 22,705 | -0,295 |
| 69 | 21 | 21,126 | 0,126 |
| 70 | 30 | 30,297 | 0,297 |
| 71 | 0 | 0,177 | 0,177 |
| 72 | 28 | 28,226 | 0,226 |
| 73 | 29 | 28,749 | -0,251 |
| 74 | 16 | 16,14 | 0,14 |
| 75 | 6 | 5,866 | -0,134 |
| 76 | 15 | 14,502 | -0,498 |
| 77 | 16 | 16,227 | 0,227 |
| 78 | 23 | 23,05 | 0,05 |
| 79 | 19 | 19,457 | 0,457 |
| 80 | 9 | 8,945 | -0,055 |
| 81 | 10 | 9,831 | -0,169 |
| 82 | 9 | 8,98 | -0,02 |
| 83 | 19 | 18,602 | -0,398 |
| 84 | 25 | 24,986 | -0,014 |
| 85 | 20 | 20,489 | 0,489 |
| 86 | 16 | 15,962 | -0,038 |
| 87 | 15 | 15,111 | 0,111 |
| 88 | 7 | 7,044 | 0,044 |
| 89 | 12 | 11,89 | -0,11 |
| 90 | 27 | 27,005 | 0,005 |

| # | Gerçek Değer | Tahmin Edilen | Hata |
|-----|--------------|---------------|--------|
| 91 | 22 | 22,162 | 0,162 |
| 92 | 22 | 21,883 | -0,117 |
| 93 | 20 | 20,147 | 0,147 |
| 94 | 17 | 17,486 | 0,486 |
| 95 | 13 | 12,9 | -0,1 |
| 96 | 18 | 18,058 | 0,058 |
| 97 | 17 | 17,468 | 0,468 |
| 98 | 4 | 3,734 | -0,266 |
| 99 | 8 | 7,896 | -0,104 |
| 100 | 29 | 29,242 | 0,242 |
| 101 | 9 | 9,056 | 0,056 |
| 102 | 31 | 31,178 | 0,178 |
| 103 | 36 | 35,69 | -0,31 |
| 104 | 20 | 19,98 | -0,02 |
| 105 | 19 | 19,133 | 0,133 |
| 106 | 14 | 14,226 | 0,226 |
| 107 | 30 | 29,984 | -0,016 |
| 108 | 20 | 19,961 | -0,039 |
| 109 | 28 | 28,204 | 0,204 |
| 110 | 15 | 15,224 | 0,224 |
| 111 | 23 | 23,031 | 0,031 |
| 112 | 3 | 2,764 | -0,236 |
| 113 | 17 | 16,991 | -0,009 |
| 114 | 18 | 17,844 | -0,156 |
| 115 | 14 | 14,085 | 0,085 |
| 116 | 7 | 7,056 | 0,056 |
| 117 | 12 | 11,904 | -0,096 |
| 118 | 18 | 18,064 | 0,064 |

Classifier

Choose

M5P -M 4.0 -num-decimal-places 4

Test options

☐ Use training set☒ Supplied test set

Set...

☐ Cross-validation

Folds

10

☐ Percentage split

%

66

More options...

(Num) TemperatureCelcius

Start

Stop

Result list (right-click for options)

19:20:36 - trees.M5P

Classifier output

| | | | |
|-----|----|--------|--------|
| 87 | 15 | 15.111 | 0.111 |
| 88 | 7 | 7.044 | 0.044 |
| 89 | 12 | 11.89 | -0.11 |
| 90 | 27 | 27.005 | 0.005 |
| 91 | 22 | 22.162 | 0.162 |
| 92 | 22 | 21.883 | -0.117 |
| 93 | 20 | 20.147 | 0.147 |
| 94 | 17 | 17.486 | 0.486 |
| 95 | 13 | 12.9 | -0.1 |
| 96 | 18 | 18.058 | 0.058 |
| 97 | 17 | 17.468 | 0.468 |
| 98 | 4 | 3.734 | -0.266 |
| 99 | 8 | 7.896 | -0.104 |
| 100 | 29 | 29.242 | 0.242 |
| 101 | 9 | 9.056 | 0.056 |
| 102 | 31 | 31.178 | 0.178 |
| 103 | 36 | 35.69 | -0.31 |
| 104 | 20 | 19.98 | -0.02 |
| 105 | 19 | 19.133 | 0.133 |
| 106 | 14 | 14.226 | 0.226 |
| 107 | 30 | 29.984 | -0.016 |
| 108 | 20 | 19.961 | -0.039 |
| 109 | 28 | 28.204 | 0.204 |
| 110 | 15 | 15.224 | 0.224 |
| 111 | 23 | 23.031 | 0.031 |
| 112 | 3 | 2.764 | -0.236 |
| 113 | 17 | 16.991 | -0.009 |
| 114 | 18 | 17.844 | -0.156 |
| 115 | 14 | 14.085 | 0.085 |
| 116 | 7 | 7.056 | 0.056 |
| 117 | 12 | 11.904 | -0.096 |
| 118 | 18 | 18.064 | 0.064 |

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.14 seconds

=== Summary ===

| | |
|-----------------------------|----------|
| Correlation coefficient | 0.9998 |
| Mean absolute error | 0.1484 |
| Root mean squared error | 0.1961 |
| Relative absolute error | 2.0277 % |
| Root relative squared error | 2.2162 % |
| Total Number of Instances | 118 |

Status

OK

Kaynaklar

1. T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü resmi internet sitesi, Erişim Tarihi: Mayıs 2020, <https://www.mgm.gov.tr/genel/ssss.aspx?s=havatahmini>