



Akarsu Ortalama Akımlarının Çeşitli Makine Öğrenme Algoritmaları Kullanılarak Tahmini: Köprüçay Örneği

Yunus Ziya KAYA^{1*}

¹Osmaniye Korkut Ata Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, İnşaat Mühendisliği Bölümü, Merkez/Osmaniye, 80100

Article Info

Research article
Received: 24/09/2024
Revision: 23/11/2024
Accepted: 24/12/2024

Keywords

Ortalama Akımlar
MLP
DVM
Random Forest
Tahmin

Makale Bilgisi

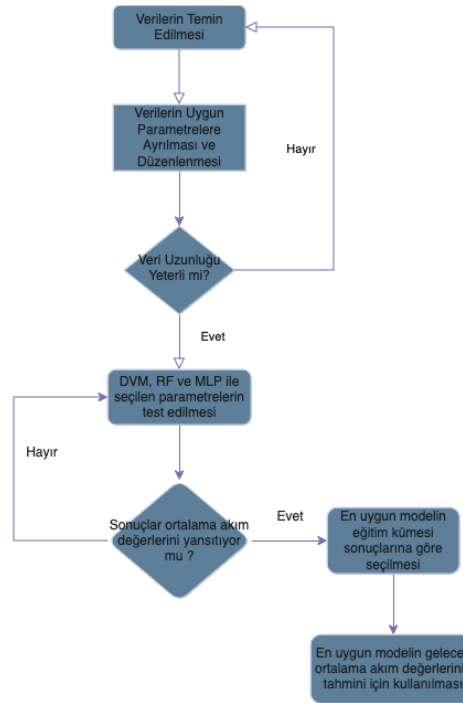
Araştırma makalesi
Başvuru: 24/09/2024
Düzeltilme: 23/11/2024
Kabul: 24/12/2024

Anahtar Kelimeler

Average Flows
MLP
SVM
RF
Prediction

Grafik Özet (Graphical/Tabular Abstract)

Köprüçay istasyonuna ait ortalama akım değerleri Multi-Layer Perceptron (MLP), Destek Vektör Makinaları (DVM) ve Random Forest (RF) makine öğrenme algoritmaları ile tahmin edilmiştir. Algoritmaların performansları karşılaştırılarak en uygun yöntem belirlenmiştir. / The average flow values of the Köprüçay station were estimated with Multi-Layer Perceptron (MLP), Support Vector Machines (SVM) and Random Forest (RF) machine learning algorithms. The performances of the algorithms were compared..



Şekil A: Köprüçay Ortalama Akımlarının Modelleme Aşamaları **Flows/ Figure A:** Modeling Processes of Köprüçay Average Stream Flow

Önemli noktalar (Highlights)

- Akarsu ortalama akımları üç popüler makine öğrenme algoritması kullanılarak tahmin edilmiştir. / River average flows were predicted by using three popular machine learning algorithms.
- Random Forest algoritması en başarılı algoritma olmuştur. / RF algorithm was detected as the best algorithm.
- Gelecek değerler RF algoritması kullanılarak tahmin edilmiştir. / Future values were predicted by using RF.

Amaç (Aim): Köprüçay ortalama akımlarını tahmin etmek. / Prediction of average stream flows.

Özgünlük (Originality): Verilen algoritmalar bu akarsu için ilk defa test edilmiştir. / The given algorithms were tested for the first time for this stream according to the knowledge of the author.

Bulgular (Results): Çalışmada RF en iyi sonucu veren algoritma olurken DVM en yetersiz sonucu vermiştir. / In this study, the RF algorithm was found to be the best, while the SVM was found to be the most insufficient.

Sonuç (Conclusion): Köprüçay ortalama akım tahmininde RF algoritmasının kullanılabilirliği gösterilmiştir. / The usability of the RF algorithm in the Köprüçay average stream prediction has been showed.



Akarsu Ortalama Akımlarının Çeşitli Makine Öğrenme Algoritmaları Kullanılarak Tahmini: Köprüçay Örneği

Yunus Ziya KAYA ^{1*}

¹Osmaniye Korkut Ata Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, İnşaat Mühendisliği Bölümü, Merkez/Osmaniye, 80100

Makale Bilgisi

Araştırma makalesi
Başvuru: 24/09/2024
Düzeltilme: 23/11/2024
Kabul: 24/12/2024

Anahtar Kelimeler

Ortalama Akımlar
MLP
DVM
Random Forest
Tahmin

Öz

Akarsu ortalama akımları havzanın su kaynaklarının yeterliliği hakkında önemli ipuçları barındırmaktadır. İklim değişikliği ile birlikte yağış ve sıcaklık gibi akarsu akımlarını doğrudan ilgilendiren parametrelerde bölgesel değişimler yaşanmaktadır. Yaşanan bu değişimler ortalama akımlarda da bölgesel farklılıklar görülmesine neden olmaktadır. Bu çalışmada Elektrik İdaresinin kayıtlarını paylaştığı Antalya ili Serik İlçesi Beşkonak Bucağında yer alan Köprüçay istasyonuna ait ortalama akımlar incelenmiştir. İstasyona ait 1957-2011 yılları arasındaki ortalama akımlar Multi-Layer Perceptron (MLP), Destek Vektör Makinaları (DVM) ve Random Forest (RF) makine öğrenme algoritmaları ile modellenmiştir. Çalışma iki kısımdan oluşmaktadır. İlk kısımda 1957-2011 yılları arasındaki veriler (55 yıllık kayıtlar) hem eğitim hem test kümesi olarak kullanılmış en uygun algoritmaya bu şekilde karar verilmiştir. İkinci kısımda algoritma seçiminden sonra kayıtları mevcut olmayan 2012-2022 yılları arasındaki ortalama akımlar tahmin edilmiştir. Modellemelerde ülkemize ait yıllık ortalama maksimum, minimum, ortalama sıcaklık ve ortalama yağış verileri girdi olarak kullanılmıştır. Analizler neticesinde en yüksek determinasyon katsayısı (R2) değeri iki numaralı Random Forest modeli için 0.947 en düşük R2 değeri ise 0.222 olarak bir numaralı DVM modeli için hesaplanmıştır. En düşük hata modelleri iki numaralı RF modeli için hesaplanmıştır. Buna göre RF2 modelinin Ortalama Mutlak Hata (OMH) değeri 4.31 Göreli Mutlak Hata (GMH) değeri ise %30.31 olarak hesaplanmıştır. Sonuç olarak Köprüçay özelinde ortalama akım tahmininde en uygun algoritmanın RF olacağı görülmüştür.

Estimation of Average Stream Flows Using Various Machine Learning Algorithms: A Case Study of Köprüçay

Article Info

Research article
Received: 24/09/2024
Revision: 23/11/2024
Accepted: 24/12/2024

Keywords

Average Flows
MLP
SVM
RF
Prediction

Abstract

The average stream flows contain important clues about the adequacy of the basin's water resources. Along with climate change, there are regional changes in parameters that directly concern stream flows, such as precipitation and temperature. These changes also cause regional differences in average flows. In this study, the average flows of the Köprüçay station located in the Beşkonak district of the Serik district of Antalya province, whose records are shared by the Electricity Administration, were examined. The average flow of the station between the years 1957 and 2011 was modeled using Multi-Layer Perceptron (MLP), Support Vector Machines (SVM), and Random Forest (RF) machine learning algorithms. The study consists of two parts. In the first part, the data between the years 1957-2011 were used as both training and test sets, and the most appropriate algorithm was decided in this way. In the second part, after the algorithm selection, the average flows between the years 2012-2022, for which records were not available, were estimated. In the models, the annual average maximum, minimum, average temperature and average precipitation data of Türkiye were used as input. As an output of the analysis, the highest coefficient of determination (R2) value was calculated as 0.947 for the Random Forest model number two, and the lowest R2 value was calculated as 0.222 for the SVM model number one. The lowest error models were calculated for RF model number two. Accordingly, the Mean Absolute Error (MAE) value of the RF2 model was calculated as 4.31, and the Relative Absolute Error (RMA) value was calculated as 30.31%. Briefly, it was seen that the most suitable algorithm for estimating the average flow in Köprüçay would be RF.

1. GİRİŞ (INTRODUCTION)

Akarsuların ortalama akımlarının değişimi iklim değişikliğine bağlı yağış ve sıcaklık parametrelerindeki değişimler ile doğrudan ilişkilidir. Ortalama akımlardaki ani artışlar akarsuyun bulunduğu havza için taşkın tehlikesi oluşturabilecekken ani düşüşler de bölgede kuraklık tehlikesine işaret edebilmektedir [1]. Akarsu ortalama akımlarının düzenli olarak ölçümü bölgenin su kaynaklarının planlanması, arz ve talep dengesinin kurulabilmesi için önemlidir. Ancak akarsu ölçümlerinin karmaşıklığı ve bazı bölgelerde yaşanan dönemsel aksaklıklar nedeniyle ülkemizdeki her akarsu için kesintisiz veri temini mümkün olamayabilmektedir. Bu durumda elde edilen veriler ile makine öğrenme (ML) teknikleri kullanılarak farklı zaman ölçekleri için debi veya ortalama akım tahminlerinin yapılabilirliği araştırma konusu olmuştur. Kızılırmak özelinde yapılan bir çalışmada Uyarlamalı Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi (ANFIS) kullanılarak iki akım gözlem istasyonu için günlük ortalama debi tahmini mekânsal yağış verileri ile yapılmıştır [2]. Sözü edilen çalışmada ANFIS bilgisayar tabanlı tahmin yaklaşımı günlük debi tahmininde başarılı bulunmuştur [2]. Başka bir çalışmada Karadeniz de bulunan Aksu deresi için yağış ve buhar basıncı parametreleri kullanılarak gelecek akım tahmininde bulunulmuştur [3]. Çalışmada Yapay Sinir Ağları (YSA) ile Klasik Regresyon yaklaşımları tercih edilmiş model çıktıları farklı istatistikler hata hesaplamaları yapılarak karşılaştırılmış, sonuç olarak modellerde gecikmeli veri seti kullanımının modelleme performansını arttırdığı vurgulanmıştır [3]. Mersin Lamas Nehri için yapılan bir ML performansı araştırmasında Destek Vektör Makinaları (DVM), YSA ve Derin Öğrenme yaklaşımları kullanılmıştır ve istatistiksel karşılaştırmalar sonucunda Lamas Nehri özelinde en uygun sonucu veren ML yaklaşımının Derin Öğrenme olduğu belirtilmiştir [4]. Aksu deresi gelecek akım tahmini için yapılan bir diğer çalışmada ise YSA, ANFIS, DVM ve farklı regresyon yöntemleri ile geliştirilen yağış akış modellerinin performansları kıyaslanmıştır, sonuç olarak Çok Değişkenli Uyarlanabilir Regresyon Eğrileri yaklaşımının diğer ML algoritmalarından daha iyi performans verdiği belirtilmiştir [5]. Birleşik Devletlerde bulunan dört farklı nehrin akım verilerinin tahmininde yağış, maksimum ve minimum hava sıcaklığı parametreleri girdi olarak seçilmiş ve DVM, YSA ve Extreme Learning Machine (ELM) algoritmalarının akım tahminlerindeki başarısı araştırılmıştır [6]. Hindistan da bulunan Narmada Nehri akım değerlerinin tahmini için yapılan bir diğer

araştırmada ise Light Gradient Boosting Machine, Random Forest, XGBoost ve CatBoost yaklaşımları kullanılmıştır [7]. Çalışmada model performansları istatistiksel hata hesaplarına dayanarak yapılmış, araştırılan algoritmalar içerisinde en uygun performansı veren algoritmanın Random Forest algoritması olduğu vurgulanmıştır [7]. İlgili literatürde farklı zaman ölçeğinde akım değerleri için de çalışmalar mevcuttur. Benzer makine öğrenme algoritmaları kullanılarak yapılan bir çalışmada Pakistan da bulunan Hunza Nehri aylık akım değerleri incelenmiştir [8]. Des Moines ve Raccoon Nehirleri için yapılan analizlerde ise Random Subspace, M5 Karar Ağacı (M5P), Random Forest ve Bagging yaklaşımları kullanılmış ve günlük debi değerleri araştırılmıştır [9]. Bu çalışmada en uygun sonuçları veren algoritmanın M5P olduğu sonucuna varılmıştır [9]. Nehir akımlarının çevresel akış, hidroelektrik üretim, su potansiyeli gibi birçok bakımdan önem taşıdığını belirten bir başka çalışmada [31], Türkiye’den iki Irak’tan iki farklı akım gözlem istasyonuna ait veriler çeşitli regresyon yaklaşımları ile tahmin edilmiştir. Çalışmada least square support vector regression (LSSVR), multivariate adaptive regression splines (MARS) and M5 Model Tree (M5-Tree) ve doğrusal regresyon yaklaşımları iki aşamalı olarak test edilmiş sonuç itibarıyla genellikle LSSVR yönteminin diğer yöntemlere kıyasla tahmin başarısının daha üstün olduğu vurgulanmıştır. Yakın zamanda yapılan bir çalışmada ise nehri akımlarının tahmininde hibrit model seçiminin önemine vurgu yapılmıştır [32]. Çalışmada GA-CatBoost yaklaşımını kombine eden bir model önerilmiş, akım tahmininde önerilen modelin performansı CatBoost, Long-Short Term Memory (LSTM) ve doğrusal regresyon yaklaşımları ile karşılaştırılmıştır. Bahsedilen çalışmada elde edilen sonuçlara göre önerilen yöntemin doğrusal olmayan nehir akımı değerlerini iyi derecede yakaladığı anlatılmıştır.

Yukarıda pek çok farklı makine öğrenme algoritmasının akarsu akımlarının tahminindeki kullanımı özetlenmiştir. Bu popüler algoritmaların hidrolojideki kullanımı ortalama akımlar ile sınırlı değildir. Hidrolojik tahmin analizlerinde bu algoritmalar evapotranspirasyon tahmini [10-12], sediment taşınımı [13] gibi pek çok spesifik olayda kullanılmıştır. Bunun yanı sıra söz edilen algoritmalar inşaat mühendisliğinin farklı alanlarında da özel kullanımlara sahiptir [14-17].

Bu çalışmada hem hidrolojik tahminlerde hem de inşaat mühendisliğinin farklı alanlarında etkinliği test edilmiş ve kanıtlanmış üç farklı makine öğrenme algoritması seçilmiştir. Bu yaklaşımlar

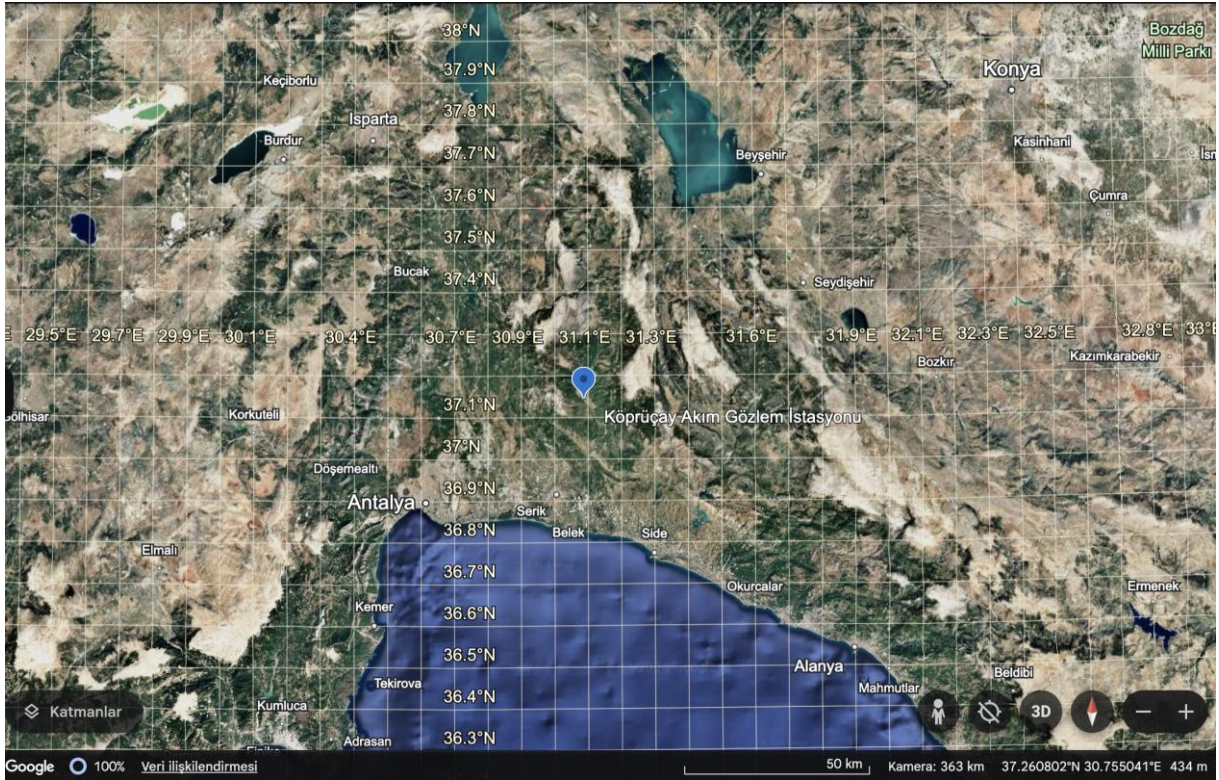
Multi-Layer Perceptron (MLP), Destek Vektör Makinaları (DVM) ve Random Forest (RF) algoritmalarıdır. Antalya il sınırları içerisinde yer alan Köprüçay akım gözlem istasyonuna ait verilerin çalışıldığı bu araştırma iki kısımdan oluşmaktadır. Çalışmanın ilk kısmında elde edilen veriler ile en uygun algoritmanın bulunması amaçlanmıştır. İkinci kısımda ise en uygun algoritma kullanılarak ortalama akım verisi mevcut olmayan (veya açık erişim olarak henüz paylaşılmayan) 2012-2022 yıllarına ait ortalama akım tahminleri yapılmıştır. Çalışmada ülkemize ait yağış, ortalama, minimum ve maksimum hava sıcaklığı parametreleri modellerde girdi olarak kullanılmıştır. Ayrıca model performanslarının iyileştirilmesi için her bir algoritma ile oluşturulan model için ötelenmiş zaman serilerinin kullanıldığı durumda incelenmiştir. Model sonuçları ortalama mutlak hata (OMH), ortalama karesel hata kökü (OKHK), göreceli mutlak hata (GMH) ve göreceli karesel hata kökü (GKHK) istatistikleri hesaplanarak karşılaştırılmıştır.

2. YÖNTEM (METHODOLOGY)

Bu bölüm dört alt başlıktan oluşmaktadır. Çalışma bölgesi, veri seti ve kullanılan algoritmalara ait açıklamalar alt başlıklar halinde verilmiştir.

2.1. Çalışma Bölgesi ve Veri Seti (Study Area and the Data Set)

Bu çalışmada kullanılan ortalama akım verileri Devlet Su İşleri web sitesinden indirilmiştir [18]. Veriler Akım Gözlem Yıllıklarından alınmıştır. Ortalama akım verilerinin Elektrik İşleri Etüt İdaresi (EİE) tarafından tutulduğu bilinmektedir. Akım verileri 2011 yılında son bulmaktadır. Yağış ve Sıcaklık parametreleri veri seti ise Dünya Bankası İklim Değişikliği Portalından temin edilmiştir [19]. Akım gözlem istasyonunun konumu Şekil 1 üzerinde gösterilmiştir.



Şekil 1. Akım Gözlem İstasyonunun Konumu (Google Earth) (Location of observation station)

Şekil 1 ile gösterilen EİE'ye ait 902 numaralı Akım Gözlem İstasyonu Antalya İli Serik İlçesinin kuzeyinde yer alan Beşkonak bucağında bulunmaktadır. Akım Gözlem İstasyonunun yaklaşık

yüksektisi 116 metredir. İstasyonda ortalama akımlar su yılı baz alınarak gerçekleştirilmektedir. Yani ortalama akım değerleri her bir su yılı için 1 Ekim-30 Eylül dönemini kapsamaktadır.

Tablo 1: Eğitim verisi istatistikleri (Training data set statistics)

Parametre	Veri Uzunluğu	Eksik Veri	Min.	Maks.	Ortalama	Standart	
						Sapma	Çarpıklık Katsayısı
Tmin.	1957-2011	-	4.04	7.36	5.44	0.62	0.36
Tmaks.	1957-2011	-	15.15	18.99	16.88	0.73	0.35
Tort.	1957-2011	-	9.58	13.15	11.14	0.66	0.41
Yağış	1957-2011	-	456.42	742.32	596.13	67.38	0.11
Ortalama Debi	1957-2011	1979	50.00	120.00	80.88	17.66	0.22

Destek Vektör Makinaları, Multi-Layer Perceptron ve Random Forest modellerinde kullanılan eğitim verilerine ait minimum, maksimum, ortalama, standart sapma ve çarpıklık katsayısı istatistikleri Tablo 1 ile verilmiştir. Bu istatistiksel veriler kullanılan veri setinin kendi içerisindeki dağılımını

göstermek için verilmiştir. Ortalama akım verisi olarak sadece 1979 yılına ait veri mevcut değildir. Bu veri bir önceki ve bir sonraki yıllara ait ortalama akım verilerinin basit ortalaması alınarak tamamlanmıştır.

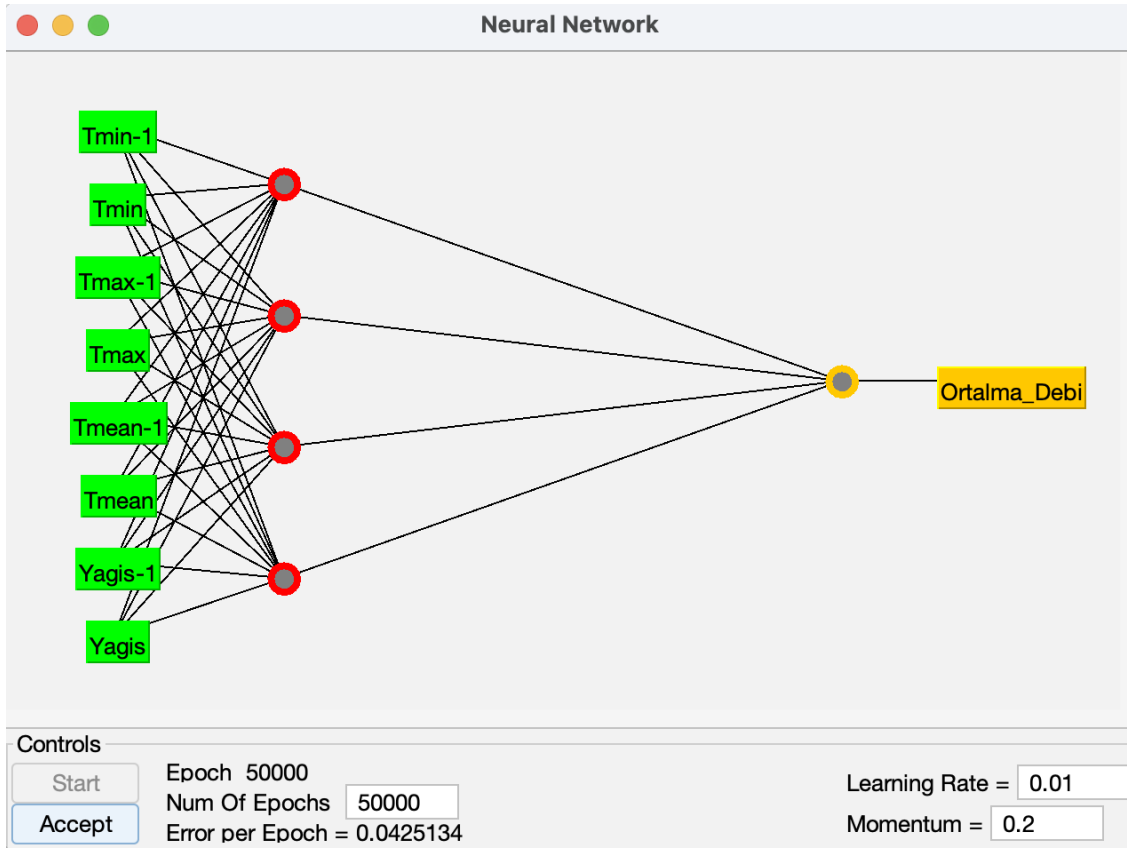
Tablo 2: Test verisi istatistikleri (Test data set statistics)

Parametre	Veri Uzunluğu	Eksik Veri	Min.	Maks.	Ortalama	Standart	
						Sapma	Çarpıklık Katsayısı
Tmin.	2012-2022	-	5.89	7.40	6.44	0.45	0.78
Tmaks.	2012-2022	-	17.50	18.86	18.15	0.42	0.15
Tort.	2012-2022	-	11.72	13.11	12.28	0.42	0.56
Yağış	2012-2022	-	518.49	711.00	604.87	68.85	0.29
Ortalama Akım	-	-	-	-	-	-	-

Tablo 2 ile çalışmanın ikinci kısmını oluşturan gelecek ortalama akım tahmininde kullanılan veri setine ait istatistikler verilmiştir. Tablo 2 de görüleceği üzere ilgili yıllara ait ortalama akım değerleri mevcut olmadığından bu veriler ile ilgili istatistikler hesaplanmamıştır. Özetle, en uygun modelin seçiminde kullanılan eğitim kümesi 55 yıllık ortalama akım değerlerini içermektedir. Kayıtların mevcut olmadığı 2012 ve sonrasında ise 11 yıllık hava sıcaklığı ve yağış verileri kullanılarak ortalama akım tahmini yapılmıştır. Eğitim amacıyla kullanılan veri setinde minimum, maksimum, ortalama hava sıcaklığı değerleri, yağış ve ortalama akım değerleri yer almakta iken test amacıyla kullanılan veri setinde eğitim kümesinden farklı olarak ortalama akım veriler mevcut değildir.

2.2. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron bazı çalışmalarda dilimize çok katmanlı algılayıcılar olarak çevrilmiştir [20, 21]. Bu çalışmada yapay sinir ağlarının özel bir uygulaması olan MLP orijinal ismi ile anılacaktır. MLP yaklaşımı eğitim için Bayesian regülarizasyonunu kullanır. Bayesian regülarizasyonu ise Levenberg-Marquardt algoritması ile ağırlık ve bias değerlerini optimize eder [22]. MLP yaklaşımında ilk adım girdi parametrelerinin gizli katmana bağlandığı ve ağırlıklarının hesaplandığı ileri besleme adımdır. Bir sonraki adım, ağırlıkları ayarlayan ters yayılım sürecidir. Son aşama ise tahmin ve gözlem arasındaki farklara uygun olarak ağırlıkları değiştiren geri yayılım sürecidir. MLP ile ilgili daha kapsamlı bilgiler ilgili literatürden [23,24] bulunabilir.



Şekil 2. Multi-Layer Perceptron Model Yapısı (MLP model structure)

Şekil 2 ile bu çalışmada kullanılan MLP yapılarından bir tanesi verilmiştir. Verilen bu MLP yapısında bütün parametrelerin mevcut ve bir önceki yıllara ait kayıtlarının girdi olarak kullanıldığı görülmektedir. Çıktı parametresi ise ortalama debi değeridir. Şekil 2 ML analizlerinin yapıldığı WEKA yazılımı ile oluşturulmuştur.

2.3. Destek Vektör Makinaları (Support Vector Machines)

Sınıflama ve regresyon konularındaki başarısı destek vektör makinalarını makine öğrenme alanındaki en popüler algoritmalarından biri yapmıştır. Cortes ve Vapnik tarafından 1995 yılında tanıtılan DVM pek çok farklı doğrusal olmayan sınıflama ya da tahmin çalışmasında kullanılmıştır [25]. Temelde bu ML yaklaşımı farklı elemanlardan oluşan bir kümedeki elemanları en doğru şekilde ayırt etmek için destek vektörleri kullanmaya dayanan bir yaklaşımdır. En uygun sınıflamayı yapmak amacıyla hiper düzlem veya karar sınırlarının belirlenmesi gerekir [26]. Ancak, destek vektörleri kullanılarak doğrusal olmayan bir veri kümesinde doğrusal bir hiper düzlem çizilemez. Bu dezavantajı çözmek için Kernel fonksiyonları kullanılır. Kernel metodu DVM tahmin ifadesi denklem 1 ile verilmiştir.

WEKA yazılımı Java programlama dilini kullanan Waikato Üniversitesinde geliştirilmiş açık kaynak kodlu bir yazılımdır. MLP modellemelerinde devir sayısı (epochs) 50000 olarak belirlenmiştir. Ayrıca öğrenme oranı (learning rate) 0.01 ve momentum değeri 0.2 olarak alınmıştır.

$$y = (K_{xi} \cdot W_{jk}) + \quad (1)$$

Denklem 1 ile verilen ifade de “b” bias terimini, “ W_{jk} ” parametre ağırlığını, “ K_{xi} ” girdi parametrelerini çok boyutlu bir düzleme aktaran doğrusal olmayan fonksiyonu tanımlamaktadır. Genel olarak bir DVM yapısı girdiler, Kernel fonksiyonu (Çekirdek Fonksiyonu olarak da adlandırılmaktadır) ve çıktılar olmak üzere üç katmandan oluşmaktadır. Belli başlı Kernel fonksiyonları Doğrusal, PolyKernel ve Radial tabanlı Kernel şeklindedir.

2.4. Random Forest

Karar ağaçları algoritmaları makine öğrenme algoritmalarındandır. Random Forest (RF) algoritması regresyon ağaçları ve kanıtsal özellikleri (evidential features) ayarlayarak tahmin modelleri oluşturur [27,28]. RF yaklaşımında araştırılan parametrenin karmaşıklığına bağlı olarak çok fazla karar ağacı oluşturulur ve çıktı

olarak karar ağaçlarının çıktılarının ortalaması alınır [29]. Karar ağaçları parametrik olmayan yaklaşımlardır ve sınıf ayırımları ve yoğunlukları için herhangi bir varsayımda bulunmaz. Bu istatistiksel karar ağaçlarında tahmin sürecinin karmaşıklığına bağlı olarak öğrenme süreci devam ettiği sürece ağaç büyür. RF tekniğinde karar ağaç sayısının artırılmasının genelleme hata oranını yakınsadığı ve bu sayede aşırı öğrenmenin bir sorun olmayacağı belirtilmektedir [30].

2.5. İş Akışı ve Hata Hesaplamaları (Flow Chart and Error Calculations)

Çalışmada bölge seçimi yapıldıktan sonra yeterli uzunluğa sahip veri (kesintisiz ve eksik veri içermeyen veri setine sahip akım gözlem istasyonu) DSİ'nin resmi websitesinden araştırılarak bulunmuştur. Akım verilerine ilaveten aynı yıllara ait ülkemizin minimum, maksimum ve ortalama hava sıcaklığı ile yağış verileri de Dünya Bankası İklim Değişikliği Portalından (Climate Change Portal) bulunarak zaman serileri halinde veriler düzenlenmiştir. Sonrasında en uygun model seçimine ve modelin kullanımına kadar olan süreci içeren iş akışı şeması Şekil 3 ile verilmiştir.

DVM analizlerinde çekirdek fonksiyonu olarak PolyKernel çekirdek fonksiyonu seçilmiştir. MLP analizlerinde öğrenme algoritması devir sayısı 50000 alınmıştır. MLP analizlerinde öğrenme oranı (learning rate) ve momentum değerleri deneme yanılma yaklaşımı ile tespit edilmiştir.

Model sonuçlarının değerlendirilmesinde kullanılan ortalama mutlak hata (OMH), ortalama karesel hata kökü (OKHK), göreceli mutlak hata (GMH) ve göreceli karesel hata kökü (GKHK) istatistiklerinin hesaplanması denklem 2-5 ile verilmiştir.

$$OMH = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (2)$$

$$OKHK = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (3)$$

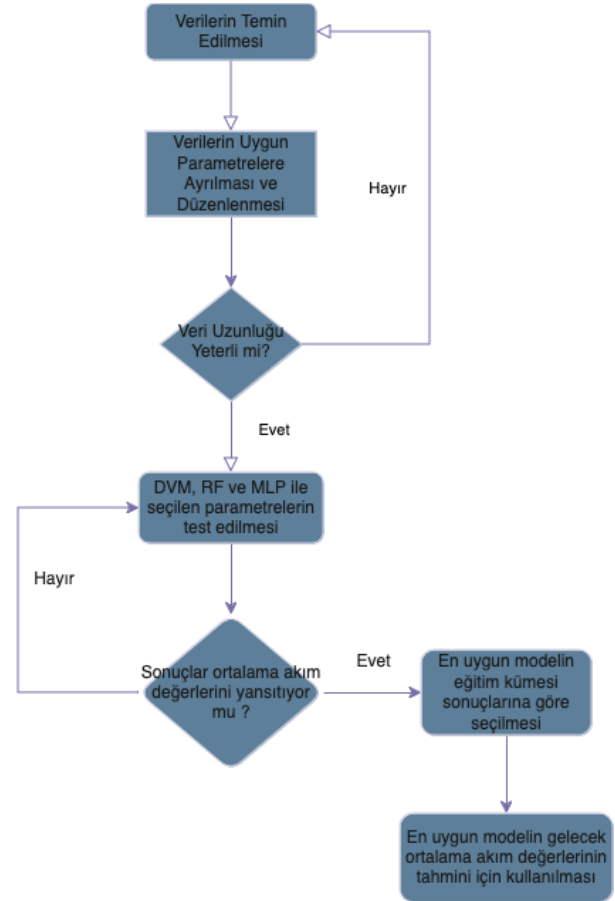
$$GMH = \frac{[\sum_{i=1}^n (y_i - x_i)^2]^{1/2}}{[\sum_{i=1}^n x_i^2]^{1/2}} \quad (4)$$

3. SONUÇLAR (RESULTS)

Bu çalışmada MLP, DVM ve RF teknikleri kullanılarak Köprüçay Akım Gözlem İstasyonunun verileri tahmin edilmiştir. Çalışmanın ilk kısmında yapılan analizlerde hangi yöntemin Köprüçay özelinde en uygun sonuçları verdiği araştırılmıştır. Bu kısımda toplamda altı adet analiz yapılmıştır. Modeller yukarıda belirtilen ML algoritmaları ve farklı girdi parametreleri kullanılarak

$$GKHK = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - \frac{1}{n} \sum_{i=1}^n x_i)^2}} \quad (5)$$

Denklem 2-5 ile verilen ifadelerde “n” veri seti uzunluğunu, “ x_i ” i’inci ortalama akım kayıt değerini, “ y_i ” ise i’inci ortalama akım tahmin değerini ifade etmektedir. Bilindiği üzere verilen ifadelerle hesaplanan istatistiksel hata değerleri düştükçe model performanslarının arttığı söylenebilir.



Şekil 3. Ortalama akım tahmini iş akışı şeması

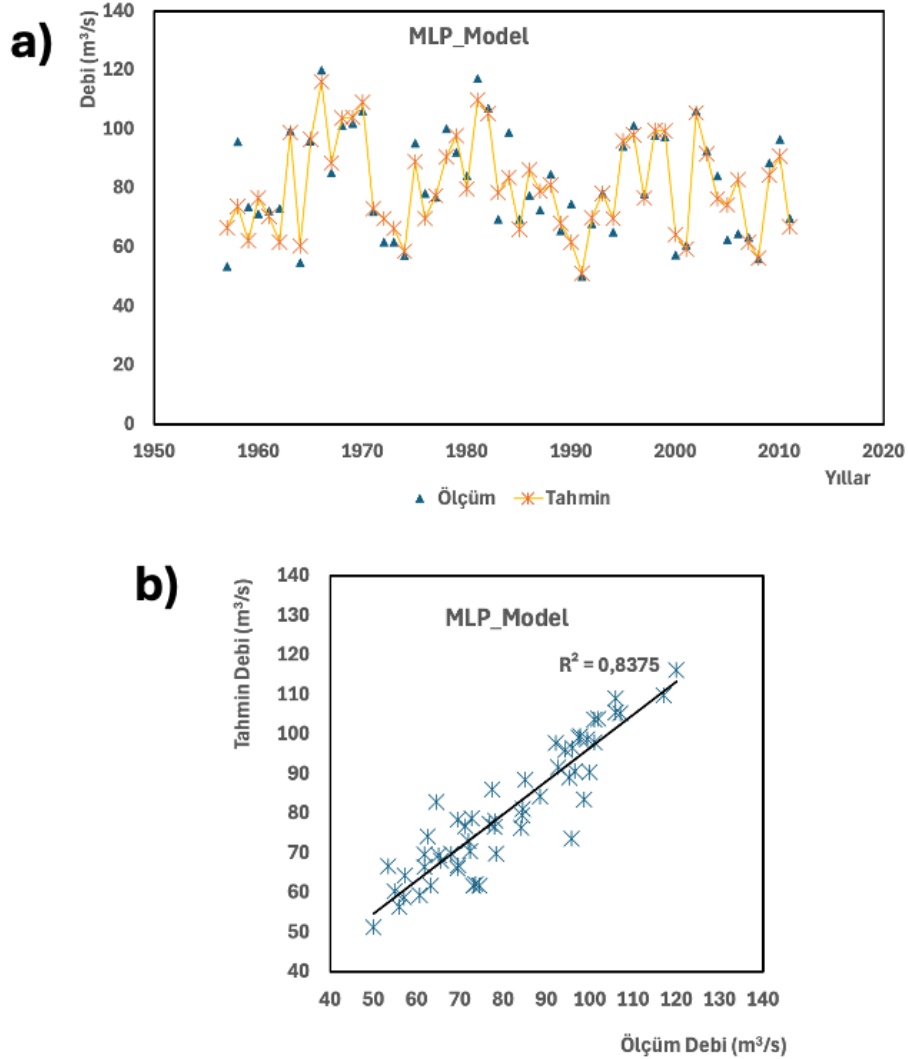
(Average flow prediction work flow)

oluşturulmuştur. Model numaraları ve girdi parametreleri Tablo 3 ile verilmiştir.

Şekil 4 ile MLP₂ model sonuçlarına ait dağılım ve saçılım grafikleri verilmiştir. MLP analizlerinde diğer bütün analizlerde olduğu gibi veriler eğitim kümesi üzerinde test edilmiştir. Bu kısımda yeni bir test kümesi kullanılmasının nedeni yeterli veri uzunluğu olmamasıdır. MLP₂ modelinde yağış ve

hava sıcaklığı parametreleri ve bu parametrelerin önceki yıllara ait kayıtları girdi olarak kullanılmıştır. Önceki yıllara ait verilerin MLP

modelinde kullanımı model performansını önemli ölçüde arttırmıştır.

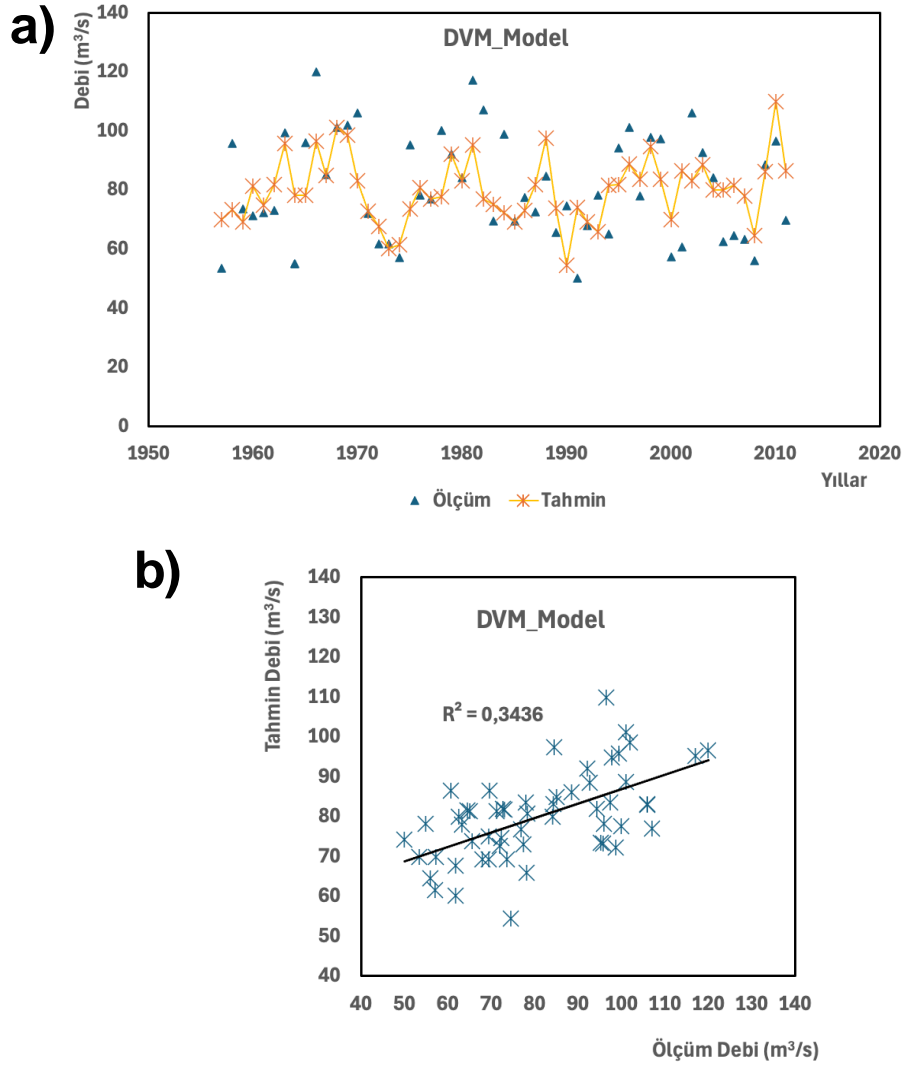


Şekil 4.a) MLP model dağılım grafiği **b)** MLP saçılım grafiği (MLP model distribution and scatter plots)

Şekil 4a ile verilen dağılım grafiğinde tahmin değerlerinin ölçüm değerlerine oldukça yakın olduğu görülmektedir. MLP modeli için saçılım grafiği üzerinde verildiği üzere determinasyon katsayısı 0.8375 olarak hesaplanmıştır. Hesaplanan determinasyon katsayısı eğitim kümesi için yapılan bu analizde tahmin ve ölçüm değerleri arasında yüksek bir ilişki olduğunu ortaya koymaktadır.

MLP₂ modelinde tahmin ve ölçüm değerleri dağılım grafiği üzerinden okunduğunda önemli ölçüde örtüşmeler olduğu anlaşılmaktadır. Ancak saçılım grafiği etrafındaki dağılım tahmin değerlerinde çok da düşük olmayan hatalara işaret etmektedir.

Şekil 5 ile DVM₂ model sonuçları dağılım ve saçılım grafikleri ile verilmiştir. DVM₂ modeli de MLP₂ modelinde olduğu gibi yağış ve sıcaklık parametreleri ile bir önceki yıllara ait kayıtların girdi olarak kullanılması ile oluşturulmuştur. Bu model de elde edilen sonuçlar kullanılabilir düzeyde anlamlı bulunmamıştır. Şekil 5 ile verilen dağılım grafiğinden tahmin ve ölçüm değerleri arasındaki uyumsuzluk anlaşılmaktadır. Ayrıca bu model için determinasyon katsayısı 0.3436 olarak hesaplanmıştır. Bu değer tahmin ve ölçüm değerleri arasındaki yakınsamanın düşük olduğunu göstermektedir. DVM₂ modeli ötelenmiş zaman serilerinin kullanıldığı DVM modeli olmasına rağmen sonuçlar kullanılabilir düzeyde değildir

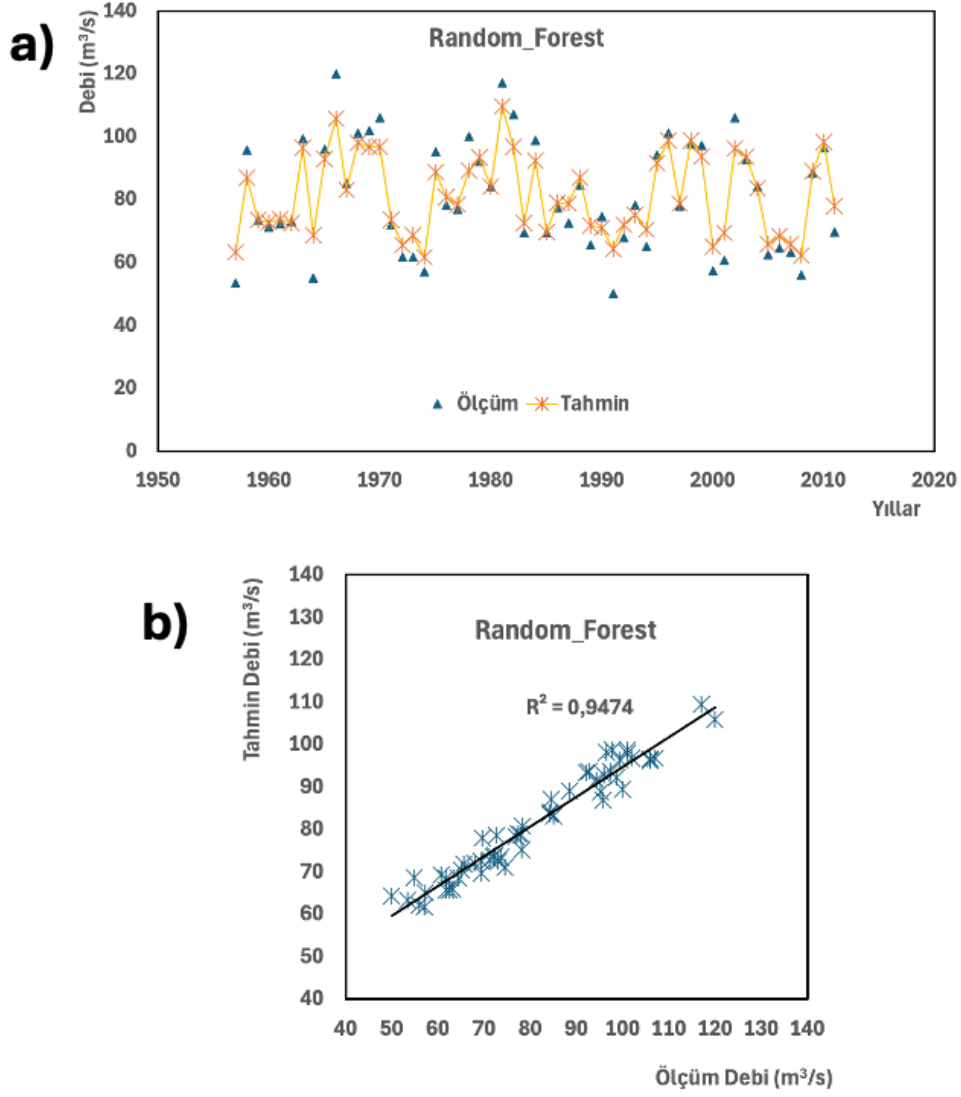


Şekil 5: a) DVM model dağılım grafiği b) DVM saçılım grafiği (SVM model distribution and scatter plots)

Şekil 6 ile iki numaralı Random Forest yaklaşımına ait sonuçlar verilmiştir. Bu modelde de önceki modellerde olduğu gibi önceki yıllara ait ötelenmiş veriler girdi olarak kullanılmıştır. RF ile yapılan analizlerde eğitim verisi için en uygun sonuçlar elde edilmiştir. RF₂ dağılım sonuçlarının verildiği Şekil 6'da tahmin ve ölçüm değerleri arasındaki uyum görülmektedir. Ayrıca Şekil 6'da verilen saçılım grafiğinden determinasyon katsayısının 0,9474 olarak hesaplandığı ve saçılımın yakınsaması görülmektedir. Bu modelde hesaplanan determinasyon katsayısı bütün modellerde hesaplanan en yüksek determinasyon katsayısı olmuştur. Oluşturulan bütün modellere ait korelasyon katsayıları ve hata hesapları Tablo 3 ile verilmiştir. Tablo 3'te girdi parametreleri kısmında "-1" ile ifade edilen girdi parametreleri bir zaman adımı önceki kayıtları ifade etmektedir. Kullanılan

girdi parametreleri ve ortalama akım değerleri yıllık verilerden olduğundan "-1" ile ifade edilen girdi parametreleri bir yıl önceki kayıtlar kullanılarak oluşturulan zaman serilerinin yeni bir parametre olarak kullanılmasını göstermektedir.

Tablo 3 ile verilen girdi parametreleri her bir makine öğrenme algoritması için aynı olacak şekilde seçilmiştir. Modeller öncelikle ötelenmiş zaman serilerinin olmadığı durum için test edilmiştir. Ötelenmiş zaman serilerinin olmadığı durum için hiçbir algoritmanın sonuçları yeterince iyi bulunmadığından sonuç grafikleri sadece ötelenmiş serilerin eklenmesi ile oluşturulan modeller için verilmiştir. Oluşturulan modellerde ortalama akım parametresi çıktı parametresi olduğundan bu parametre için ötelenmiş zaman serisi tercih edilmemiştir.



Şekil 6: a) RF model dağılım grafiği b) RF saçılım grafiği (RF model distribution and scatter plots)

Tablo 3 incelendiğinde en yüksek korelasyon katsayısının RF₂ modeli için elde edildiği görülecektir. RF₂ modelinde yağış ve sıcaklık parametreleri ve bunlara ait ötelenmiş zaman serileri birlikte girdi olarak kullanılmıştır. En düşük korelasyon katsayısı ötelenmiş zaman serilerinin kullanılmadığı DVM₁ modeli için hesaplanmıştır. Her üç ML yaklaşımında da ötelenmiş zaman serilerinin korelasyon katsayısını önemli ölçüde arttırdığı saptanmıştır. En düşük OMH, OKHK,

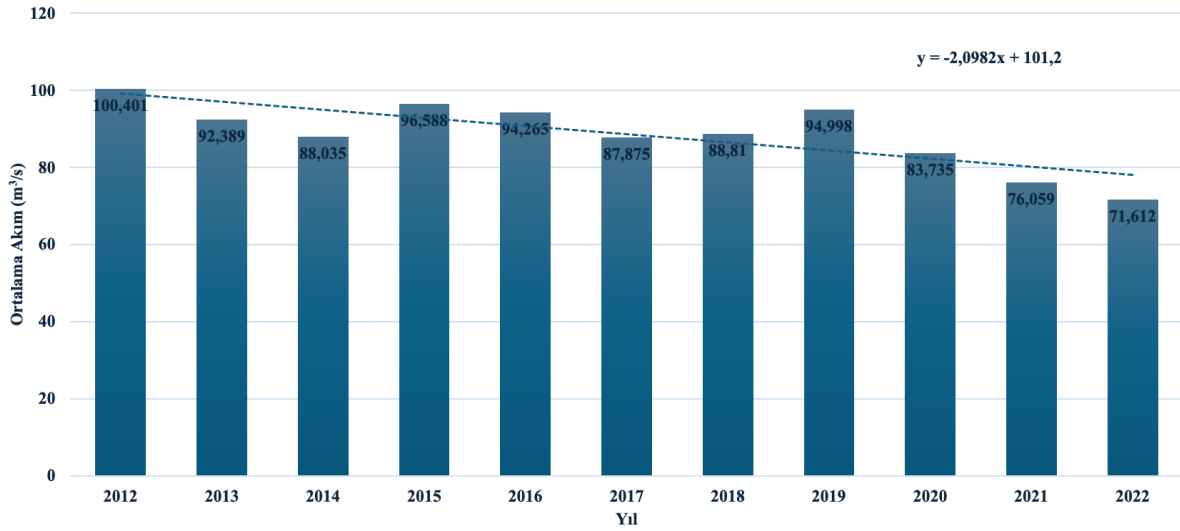
GMH ve GKHK değerleri RF₂ modeline aittir. En yüksek hata oranları ise DM₁ modeli için bulunmuştur. Hesaplanan hata değerleri ve korelasyon katsayıları RF₂ modelinin seçilen modeller içerisinde en uygun performansı verdiğini ortaya koymuştur. Tablo 3 ile verilen MLP modelleri incelendiğinde ise MLP₂ model sonuçlarının tahmin süreçlerinde kullanılabilir düzeyde anlamlı olduğunu göstermiştir.

Tablo 3. Model sonuçları ve performans analizleri (Models results and performance analyses)

Algoritma	Girdi Parametreleri	Korelasyon Katsayısı	OMH	OKHK	GMH (%)	GKHK (%)
DVM₁	Tmin, Tmax, Tort., Yağış,	0.4714	12.43	15.49	81.47	88.54
DVM₂	Tmin, Tmin-1, Tmax, Tmax-1, Tort., Tort.-1, Yağış, Yağış-1	0.5861	11.28	14.21	73.91	81.25
MLP₁	Tmin, Tmax, Tort., Yağış,	0.6207	11.11	13.72	72.79	78.42
MLP₂	Tmin, Tmin-1, Tmax, Tmax-1, Tort., Tmean-1, Yağış, Yağış-1	0.9152	5.22	7.00	34.23	40.35
RF₁	Tmin, Tmax, Tort., Yağış,	0.9536	5.31	6.91	34.78	39.49
RF₂	Tmin, Tmin-1, Tmax, Tmax-1, Tort., Tort.-1, Yağış, Yağış-1	0.9733	4.62	5.95	30.31	34.00

Çalışmanın ikinci kısmında en uygun sonuçları veren RF₂ yaklaşımı kullanılarak sonraki yılların yağış ve sıcaklık parametreleri ile ortalama akım tahmini yapılmıştır. Yağış ve sıcaklık parametrelerine ait veriler 2022 yılında son

bulduğundan tahminler ortalama akım verilerinin bittiği 2011 yılından 2022 yılına kadar olan süre için yapılmıştır. Şekil 7 ile yıllara göre tahmin sonuçları histogram ile verilmiştir.



Şekil 7. 2012- 2022 yıllarına ait ortalama debi tahmini sonuçları (Average flows prediction results of 2012-2022)

Şekil 7’de verilen grafiğe eklenen doğrusal eğilim çizgisinin yıllar içerisinde ortalama akımda bir düşüşe işaret ettiği anlaşılmaktadır. Özellikle 2021 ve 2022 yıllarına ait tahminlerde ciddi oranda düşüşler gözlemlenmiştir. Bu sonuçlar küresel sıcaklık artışlarına bağlı değişimler nedeniyle özellikle son yıllarda kırılan sıcaklık rekorları ve değişen yağış rejimleri düşünüldüğünde anlamlı sonuçlar olarak değerlendirilmiştir.

4. DEĞERLENDİRME (CONCLUSIONS)

Akarsular buldukları bölgenin içme ve sulama suyu ihtiyacını karşılayan önemli su kaynaklarıdır. Son yıllarında değişen iklim koşulları ile birlikte akarsu debilerinde önemli değişiklikler olduğu bilinmektedir. Bu çalışmada EİE kayıtları kullanılarak Köprüçay özelinde ortalama akımlara ait değerler çeşitli makine öğrenme algoritmalarından faydalanılarak tahmin edilmeye çalışılmıştır. Yağış, minimum hava sıcaklığı, maksimum hava sıcaklığı, ortalama hava sıcaklığı parametreleri ile bu parametrelerin ötelenmiş zaman serilerini içeren 55 yıllık bir veri seti kullanılmıştır. Veri setinin ML araştırmaları için yeterli uzunluğa sahip olmadığı düşünüldüğünden DVM, RF ve MLP algoritmalarının eğitim performansları ilk etapta değerlendirmeye tabi tutulmuştur. Çalışmanın bu ilk kısmında elde edilen hata değerleri RF yaklaşımının en uygun sonuçları verdiğini göstermiştir. Çalışmanın ikinci kısmında ise RF modeli kullanılarak 2012-2022 yılları için ortalama akımlar tahmin edilmiştir. Çalışmada bölgesel hava sıcaklığı ve yağış verileri temin edilemediğinden, açık erişim ile ulaşılabilen uzun yıllara ait ülkemizin sıcaklık ve yağış verileri tercih edilmiştir ve ülkemiz özelinde iklim değişikliğinin önemli bir

göstergesi olan bu parametrelerin model sonuçları üzerinde de etkili olduğu tespit edilmiştir. 2012 yılı ve sonrası için yapılan tahmin sonuçları ortalama akımların azalma eğiliminde olduğunu göstermiştir. Benzer bir yaklaşımla havzadaki diğer akarsu ortalama akımlarının incelenebileceği ve sonuçlarının su yönetimi planları için kullanılabilirliği düşünülmektedir. Bu çalışmada kullanılan bütün makine öğrenme algoritmalarında ötelenmiş zaman serilerinin sonuçlar üzerinde olumlu etkileri olduğu görülmüştür. Çalışmada çıktı parametresi olan ortalama akım parametresine ait ötelenmiş zaman serisi kullanılmamıştır. Ötelenmiş zaman serileri kullanılmasına rağmen DVM yaklaşımının Köprüçay özelinde ortalama akım tahmininde yetersiz kaldığı anlaşılmıştır. MLP ve RF yaklaşımlarının ortalama akım tahmininde başarılı olduğu ancak hesaplanan istatistiksel hata değerlerinin düşüklüğü nedeniyle RF modelinin daha iyi sonuçlar verdiği tespit edilmiştir.

TEŞEKKÜR (ACKNOWLEDGMENTS)

Yazar akım verilerini toplayıp, düzenleyip paylaşan Elektrik İdaresi ve Devlet Su İşleri Genel Müdürlüğüne (DSİ) teşekkür eder. / The author would like to thank the Electricity Administration and the General Directorate of State Hydraulic Works (DSI) for collecting, organizing and sharing the flow data.

ETİK STANDARTLARIN BEYANI (DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan eder. /The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

YAZARLARIN KATKILARI (AUTHORS' CONTRIBUTIONS)

Yunus Ziya KAYA: Modellerin oluşturulmasında, sonuçlarının analizinde ve makalenin yazımında görev almıştır. / He generated models, analyzed the results and performed the writing process.

ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

Bu çalışmada herhangi bir çıkar çatışması yoktur. / There is no conflict of interest in this study.

KAYNAKLAR (REFERENCES)

[1] Çubukçu, E. A., Demir, V., & Sevimli, M. F. (2022). Estimating streamflow data with machine learning techniques. *Gazi Journal of Engineering Sciences*, 8(2), 257-272.

[2] Demir, H. N., Korkmaz, M. S., & Arıman, S. (2023). Akım Gözlem İstasyonlarında ANFIS Yöntemi ile Günlük Ortalama Debi Tahmini: Kızılırmak Örneği. *Journal of the Institute of Science and Technology*, 13(4), 2565-2582. <https://doi.org/10.21597/jist.1323346>

[3] Babacan, H. T., & Saka, F. (2022). Makine Öğrenmesi İle Aksu Deresi'nde Akış Tahmin Modeli Geliştirilmesi. *Türk Hidrolik Dergisi*, 6(1), 1-11.

[4] Çubukçu, E. A., Demir, V., & Sevimli, M. F. (2022). Akım Verilerinin Makine Öğrenmesi Teknikleriyle Tahmin Edilmesi. *Gazi Mühendislik Bilimleri Dergisi*, 8(2), 257-272.

[5] Babacan, H. T., Yüksek, Ö., & Saka, F. (2022). Yapay zeka ve sezgisel regresyon yöntemlerinin yağış-akış modellemesi için performans değerlendirmesi: Aksu Deresi için bir uygulama. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 11(3), 744-751. <https://doi.org/10.28948/ngumuh.1079616>

[6] Parisouj, P., Mohebzadeh, H., & Lee, T. (2020). Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zones in the United States. *Water Resources Management*, 34(13), 4113-4131.

[7] Kedam, N., Tiwari, D. K., Kumar, V., Khedher, K. M., & Salem, M. A. (2024). River stream flow prediction through advanced machine learning models for enhanced accuracy. *Results in Engineering*, 22, 102215.

[8] Hussain, D., & Khan, A. A. (2020). Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth Science Informatics*, 13(3), 939-949.

[9] Elbeltagi, A., Di Nunno, F., Kushwaha, N. L., De Marinis, G., & Granata, F. (2022). River flow rate prediction in the Des Moines watershed (Iowa, USA): A machine learning approach. *Stochastic Environmental Research and Risk Assessment*, 36(11), 3835-3855.

[10] Kaya, Y. Z., Zelenakova, M., Üneş, F., Demirci, M., Hlavata, H., & Mesaros, P. (2021). Estimation of daily evapotranspiration in Košice City (Slovakia) using several soft computing techniques. *Theoretical and Applied Climatology*, 144, 287-298.

[11] Amani, S., & Shafizadeh-Moghadam, H. (2023). A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. *Agricultural Water Management*, 284, 108324.

[12] Sharafi, S., & Mohammadi Ghaleni, M. (2024). Revealing accuracy in climate dynamics: enhancing evapotranspiration estimation using advanced quantile regression and machine learning models. *Applied Water Science*, 14(7), 162.

[13] Üneş, F., Taşar, B., Demirci, M., Zelenakova, M., Kaya, Y. Z., & Varçin, H. (2021). Daily suspended sediment prediction using seasonal time series and artificial intelligence techniques. *Rocznik Ochrona Środowiska*, 23.

[14] Avcı, Y., & Ekmen, A. B. (2023, October). Artificial intelligence assisted optimization of rammed aggregate pier supported raft foundation systems based on parametric three-dimensional finite element analysis. In *Structures* (Vol. 56, p. 105031). Elsevier.

[15] Altay, G., Kayadelen, C., & Kara, M. (2024). Model selection for prediction of strong ground motion peaks in Türkiye. *Natural Hazards*, 120(2), 1443-1461.

[16] Karatas, I. & Budak, A. (2024), "Development and comparative of a new meta-ensemble machine learning model in predicting construction labor productivity", *Engineering, Construction and Architectural Management*, Vol. 31 No. 3, pp. 1123-1144. <https://doi.org/10.1108/ECAM-08-2021-0692>

- [17] Murad, Y., Hajar, H. A., & Azim, I. (Eds.). (2022). Machine learning applications in Civil engineering (Vol. 16648714). Frontiers Media SA.
- [18] DSİ, (2024). Url: <https://www.dsi.gov.tr/Sayfa/Detay/744> Erişim Tarihi: 17.09.2024.
- [19] World Bank, Climate Change Knowledge Portal (2024). URL: <https://climateknowledgeportal.worldbank.org/>. Erişim Tarihi: 17.08.2024.
- [20] Moralı, İ. A., & Aygün, F. F. (2007). Çok Katmanlı Algılayıcı Ve Geriye Yayılım Algoritması İle Konuşmacı Ayırt Etme. Akademik Bilişim, 7, 57-62.
- [21] İşeri, İ., & Arıman, S. (2019). Sedimandaki Ağır Metal Konsantrasyonunun Çoklu Değişken Regresyon Modelleri ve Çok Katmanlı Algılayıcı Ağ Modeli ile Tahmini. Avrupa Bilim ve Teknoloji Dergisi, 389-397.
- [22] Toprak ZF, Cigizoglu HK (2008) Predicting longitudinal dispersion coefficient in natural streams by artificial intelligence methods. Hydrol. Process. 22:4106–4129. <https://doi.org/10.1002/hyp.7012>
- [23] Bishop C (1995) Neural networks for pattern recognition. Oxford:University Press
- [24] Haykin S (1999) Neural networks: a comprehensive. Pearson Education.13:409–412. <https://doi.org/10.1017/S0269888998214044>
- [25] Cortes, C., Vapnik, V., 1995. Support-Vector Cortes, C., & Vapnik, V..Support-vector networks. Machine Learning, 20(3), 273–297. doi:<https://doi.org/10.1023/A:1022627411411>
- [26] Turan, A. K., & Polat, H. (2024). Yarı denetimli makine öğrenmesi yöntemini kullanarak müzik türlerinin tespiti. Gazi University Journal of Science Part C: Design and Technology, 1-1.
- [27] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
- [28] Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS journal of photogrammetry and remote sensing, 67, 93-104.
- [29] Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J. C. W., Van den Bossche, P., ... & Omar, N. (2018). Random forest regression for online capacity estimation of lithium-ion batteries. Applied energy, 232, 197-210.
- [30]. Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804-818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>
- [31] Yaseen, Z. M., Kisi, O., & Demir, V. (2016). Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. Water resources management, 30, 4125-4151.
- [32] Ahmadianfar, I., Demir, V., Heddami, S., Al-Areeq, A. M., Abba, S. I., Tan, M. L., ... & Yaseen, Z. M. (2023). Daily scale streamflow forecasting based-hybrid gradient boosting machine learning model.