



Destek Vektör Makineleri ve M5 Karar Ağacı Yöntemleri Kullanılarak Yağış-Akış İlişkisinin Tahmini

Mustafa DEMİRCİ *

İskenderun Teknik Üniversitesi, İnşaat Mühendisliği Bölümü, İskenderun-Hatay
mustafa.demirci@iste.edu.tr ORCID: 0000-0002-3249-2586

Geliş: 11.02.2019, Revizyon: 13.03.2019, Kabul Tarihi: 01.04.2019

Öz

Havza yönetimi ve afetlerin engellenmesi, su kaynaklarının daha verimli kullanılması ve su yapılarının inşasının planlaması amacı ile yağış ve akış verilerinin tahmini büyük önem taşımaktadır. Bu çalışmada Amerika Birleşik Devletleri Waltham Massachusetts'de yer alan Stony Brook rezervuarını besleyen Stony Brook nehrindeki 731 günlük yağış, akış ve sıcaklık bilgilerini içeren veriler kullanılarak modeller oluşturulmuştur. Bu veriler Destek Vektör Makineleri (SVM) ve M5 Karar Ağacı (M5T) yöntemlerinde girdi olarak kullanılmış ve yağış-akış ilişkisi tahmin edilmiştir. Her iki yöntemle elde edilen sonuçlar gerçek ölçüm sonuçları ile karşılaştırılmaları yapılmıştır. Bunun sonucunda M5 Karar Ağacı (M5T) modellerinin akış tahmininde daha iyi performansa sahip olduğu görülmüştür.

Anahtar Kelimeler: Yağış-Akış; Tahmin; Destek Vektör Makineleri; M5 Karar Ağacı

* Yazışmaların yapılacağı yazar

Giriş

Yağış ve akış hidroloji bilimi için en önemli parametrelerdendir. Bir havzaya düşen yağışın belli aşamalardan geçip en son akışa dönmesi ve bunun miktarının doğru analiz edilmesi taşkın kontrolü, su temini ve hidroelektrik santralleri ile enerji üretimi gibi su yapıları tesislerinin inşası için yapılacak planlamalarda doğru bir projelendirme ve uygulama açısından mühendisler için büyük bir önem taşımaktadır. Bu analizler yapılırken günlük, aylık ve yıllık yağış verileri kullanılabilir. Veriler ne kadar geniş bir aralıkta ve istikrarlı olursa sonuçlar da o kadar tutarlı olur. Yağış akış ilişkisinin belirlenmesinde genel olarak iki yöntemden bahsedilebilir. Bunlar parametrik ve sistem yaklaşımı olarak bilinen Kara Kutu modelleridir. Parametrik modeller yağışın akışa dönüşme sürecini bütün ayrıntıları ile ele alır. Havzayı sızma, buharlaşma, yüzeysel ve yer altı akışı şeklinde bütün parametrelerine ayırır. Kara Kutu modellerinde ise havzada gerçekleşen fiziksel olaylar dikkate alınmaz. Havzaya sadece yağış akışa dönüştüren matematiksel bir fonksiyon gözüyle bakılır.

Yapay Zeka yöntemleri de kara kutu modelleri olarak adlandırılmakta ve son yıllarda hidrolik ve su yapıları planlamalarında sıkça kullanılmaktadır. (Demirci vd, 2015a, 2015b, 2016, 2017, 2018a, 2018b; Kaya vd., 2016, 2018; Tasar vd, 2017, 2018; Demirci and Baltacı, 2013; Unes, 2010a, 2010b; Unes ve Demirci, 2015; Unes vd, 2015, 2017a, 2018a, 2018b) . Kara kutu modellerinden Yapay Sinir Ağları (YSA) öğrenebilme ve genelleme özellikleri nedeniyle karmaşık problemleri başarı ile çözebilme yeteneğini ortaya koymaktadır (Ergezer vd, 2003). Yapay sinir ağlarını ile ilgili ilk modellemenin temelleri McCulloch ve Pitts'in, (1943) yılında yayınladıkları bir makaleyle atılmıştır. Çalışmalarında elektrik devreleri üzerinde basit yapay sinirler modellemiştir. Günümüzde YSA'nın su kaynaklarında sıkça karşılaşılan değişik problemlere uygulanması ile ilgili çok sayıda çalışma bulunmaktadır. YSA yaklaşımının yağış-akış ilişkisini iyi temsil ettiği gösterilmiştir (Hsu vd, 1995; Mason vd, 1996; Minns ve Hall, 1996; Fernando ve Jayawardena, 1998; Unes vd, 2018c).

Hamidi ve Demirci (2006) Botan çayında ve Hamidi ve Kayaalp (2008), Dicle nehrinde YSA ve Regresyon yöntemlerini kullanarak yağış, akış ve sediment ilişkilerinde çok sayıda modeller üretmişlerdir. Turhan (2012) yaptığı çalışmasında YSA yöntemi ile Seyhan Havzası'nın yağış-akış ilişkisini modellemiştir.

Destek Vektör Makineleri (SVM), Vapnik ve diğer araştırmacılar tarafından geliştirilen makine öğrenme algoritmalarıdır (Vapnik, 1995, Cortes ve Vapnik, 1995, Vapnik vd, 1997). Bu yaklaşım ilk başta sensör verilerini yorumlamanın bir yolu olarak geliştirildi ve iki katmanlı bir yapı içermektedir. İlk katman giriş değişken serisini içeren destek vektörleri üzerindeki ağırlıksız doğrusal olmayan bir çekirdektir ve ikincisi çekirdek çıktılarının ağırlıklı bir toplamıdır. Birçok durumda, destek vektörleri ve uygun çekirdek filtreleri belirlendikten sonra SVM'ler, YSA yöntemlerine göre daha verimli olabilir. Birçok araştırmacı, SVM'leri kullanarak hidrolojik problemleri çözme olasılığını araştırmıştır. Dibike vd, (2001), YSA modellerine kıyasla yağış-akış modellemesi için SVM'leri araştırmıştır. Bray ve Han (2004), yağış modellemesi için uygun bir model yapısının ve ilgili parametrelerin belirlenmesindeki zorlukların altını çiz(diler)miştir. Chen vd, (2010) Hanjiang havzasında günlük yağışları tahmin etmek için SVM'leri kullanmıştır. SVM tekniği son zamanlarda ormancılık problemlerini çözmeye Nieto vd, (2012), güneş ışınlarını belirlemede ve günlük yağış akışını tahmin etmede başarılı olmuştur. Son zamanlarda, SVM'ler özellikle yağış-akış (Hosseini ve Mahjouri, 2016) ve su rezervuarlarını(Khan ve Coulibaly 2006; Asefa vd, 2005; Khalil vd, 2006) tahmin etmede yaygın olarak kullanılmıştır.

Yapay zeka yöntemlerinden bir tanesi olan M5 karar ağacı yöntemi (M5T), sınıflandırma ve regresyon problemlerine uygulanabilen, etkili, parametrik olmayan, yoğun hesaplama gerektiren bir yöntemdir. Birçok araştırmacı çalışmalarında yağış-akış tahmininde M5T'yi kullanmışlardır (Rao vd, 2005; Bathettchery ve Solomatine, 2005; Solomatine ve Xue, 2004). Sattari vd, (2013), M5T karar ağacını yöntemini kullanarak, Türkiye, Ankara, Sohu

Nehri'ndeki günlük akarsu akışını potansiyelini araştırmışlar ve M5T sonuçları, destek vektör makinaları (SVM) ile karşılaştırılmıştır.

Bu çalışmada, ABD'nin Stony Brook Rezervuarı havzasında bulunan 01104480 numaralı gözlem istasyonuna ait 731 günlük yağış, sıcaklık ve akış verileri girdi olarak girilip yeni akış değerleri Destek Vektör Makineleri (SVM) ve M5 Karar Ağacı (M5T) yöntemleriyle belirlenmeye çalışılmış ve sonuçlar birbirleri ile karşılaştırılmıştır.

Materyal ve Yöntem

Çalışma Alanı ve Uygulama

Amerika Birleşik Devletleri'ndeki Massachusetts eyaletinde yer alan Waltham Barajı yakınlarındaki 42°21'20" Enlem ve 71°15'56" Boylamındaki Stony Brook rezervuarını besleyen Stony Brook nehri çalışma alanı olarak seçilmiştir. Bu çalışmada drenaj alanı 61.8 km² ve depolama kapasitesi 1.340.000 m³ olan Stony Brook havzası Stony Brook nehri istasyonundaki veriler kullanılmıştır.

Amerika Birleşik Devletleri Jeolojik Araştırma Kurumu'ndan (USGS) alınan 01104480 numaralı istasyon verileri 07.05.2011-07.05.2013 tarihleri arasındaki 731 günlük yağış, akış ve sıcaklık bilgilerini içermektedir. Çalışma alanının konumu Şekil 1'de gösterilmiştir.



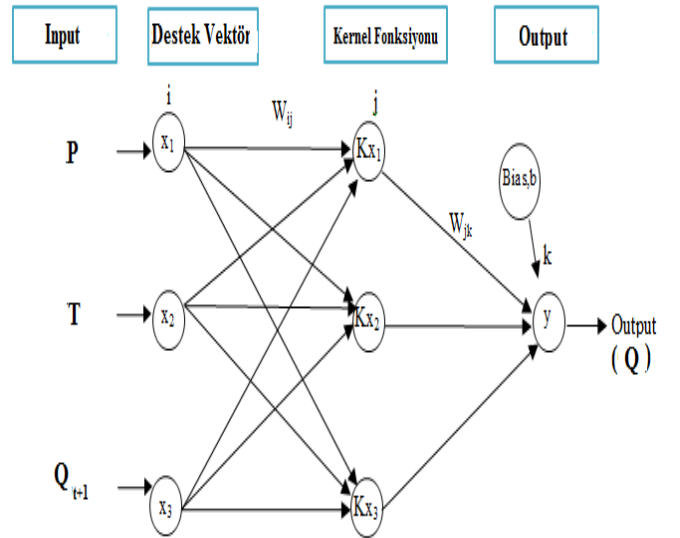
Şekil 1. Çalışma alanı (Google Maps)

Destek Vektör Makineleri (SVM)

Destek vektör makineleri (SVM), Vapnik (1995) tarafından bulunan veri odaklı araştırma alanlarında makine öğrenme yaklaşımıdır. SVM istatistiksel öğrenme teorisine dayanmaktadır. SVM temel olarak iki veri sınıfını en iyi şekilde ayırt etmek için kullanılır. Bu amaçla karar sınırları veya hiper düzlemler belirlenir. Doğrusal olmayan bir veri kümesinde, SVM'ler doğrusal bir hiper düzlem çizemez. Bu nedenle çekirdek numaraları kullanılır. Kernel yöntemi, doğrusal olmayan verilerdeki makine öğrenimini büyük ölçüde artırır. Kernel yönteminde SVM tahminçisi (y) süreci şu şekilde ifade edilebilir:

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

Denklem (1)'de b, SVM ağının bias terimidir ve W_{jk} ağırlık vektörü olarak adlandırılır. K_{xi} , giriş vektörlerini yüksek boyutlu bir özellik alanına eşleyen doğrusal olmayan bir işlevdir. Bu çalışmada kullanılan üç katmana sahip SVM modelinin yapısı Şekil 2'de verilmiştir. Bu örnek SVM modeli için çıktı değeri, girişlerdeki üç ürününün toplamına ve Lagrange çarpanlarının bağımsız kombinasyonlarına eşittir. Bu çalışmada kullanılan doğrusal olmayan çekirdek fonksiyonları, Radyal temelli çekirdek fonksiyonu ve Poli çekirdek fonksiyonlarıdır.



Şekil 2.Çalışmada kullanılan SVM-1 modelinin yapısı

Radyal Temelli Çekirdek Fonksiyonlu Destek Vektör Makineleri (SVM-RBF)

Lagrange çarpanları çıktı verileri için eğitim veri setlerinin önemini gösterir. Doğrusal olmayan radyal temelli çekirdek fonksiyonu (Hsu vd, 2003) aşağıdaki denklemle gösterilmiştir.

$$K_{xi} = e^{-\gamma \|p_i - y_i\|^2} \quad \gamma > 0 \quad \text{and} \quad i = 1, 2, 3, \dots, n \quad (2)$$

Burada γ kullanıcı tanımlı bir parametredir.

Poli Çekirdek Fonksiyonlu Destek Vektör Makineleri (SVM-PK)

Poli çekirdek fonksiyonu (Hsu vd, 2003) tarafından aşağıdaki denklemle gösterilmiştir.

$$K_{xi} = (p \cdot y + c)^d \quad i = 1, 2, 3, \dots, n \quad (3)$$

M5 Karar Ağacı Yöntemi (M5T)

Bir karar ağacı, bir bağımlı değişkenin değerinin tahminini bağımsız değişken kümesinin değerlerini kullanarak nasıl tahmin edilebileceğini gösteren bir ikili (iki yönlü bölünmüş) ağaç olarak gösterilen mantıksal bir modeldir. Temel olarak iki tür karar ağacı vardır:

(1) sınıflama ağaçları en yaygın olanıdır ve bir sayısal niteliğin değerini tahmin etmek için kullanılan bir sembolik sınıftır.

(2) regresyon ağaçları öngörmek için kullanılır. (Witten ve Frank, 2005).

Ağaçtaki her yaprak, o yapraktaki hedef değişkeni tahmin etmek için kullanılan doğrusal bir regresyon modeli içeriyorsa, o zaman bir model ağacı denir.

M5 karar ağacı algoritması orijinal olarak Quinlan (1992) tarafından geliştirildi. Bu tekniğin detaylı açıklaması Witten ve Frank (2005)'de bulunabilir. Bu tekniğin kısa bir açıklaması aşağıdaki gibidir. M5 algoritması, hedef boşluğundaki varyansı maksimuma indiren tek bir öznelik üzerinde testler kullanarak örnek boşluğunu yinelemeli olarak bölerek bir regresyon dizisi oluşturur. Standart

sapma azaltma (SDR) hesaplamak için matematiksel formül:

$$SDR = sd(T) - \sum I \frac{T_i}{T} sdIT_i \quad (4)$$

burada T düğüme ulaşan bir örnek grubunu, T_i potansiyel kümenin sonucu olan örneklerin alt kümesini ve sd standart sapmayı temsil eder. Ağaç büyütüldükten sonra, o düğüm ile ilgili verileri ve o düğümün alt testindeki testlere katılan tüm nitelikleri kullanarak her iç düğüm için doğrusal bir çoklu regresyon oluşturulur. Bundan sonra, her bir alt ağaç budama probleminin üstesinden gelmek için budama işleminde dikkate alınır. Budama, bir alt ağacın kökündeki doğrusal model için tahmini hata alt ağaç için beklenen hataya eşit veya daha küçük olduğunda oluşur. Son olarak, budama ağacının yapraklarındaki bitişik doğrusal modeller arasındaki keskin süreksizlikleri telafi etmek için yumuşatma işlemi kullanılır.

Bulgular ve Tartışma

Yağış-akış ilişkisi tahin etmek için yapılan bu çalışmada, istasyondan toplam 731 günlük veri toplanmıştır. Bu verilerin 550 adedi yani %75'i eğitim sürecinde kalan 181 adedi ise test aşamasında kullanılmıştır. Radyal Temelli Çekirdek Fonksiyonlu Destek Vektör Makineleri (SVM-RBF), Poli Çekirdek Fonksiyonlu Destek Vektör Makineleri (SVM-PK) ve M5 Karar Ağacı Yöntemleri (M5T) yöntemleri kullanılarak tahminler yapılmıştır. Her üç modelde de üçer adet değerlendirme yapılmıştır. Modellerin tahmin sonuçları R^2 (Determinasyon Katsayısı), Ortalama Mutlak Hata (OMH) ve Ortalama Karesel Hata (OKH) kriterlerine göre karşılaştırılmıştır.

SVM RBF 1, SVM PK 1 ve MP5 1 modellerinde girdi olarak Ortalama sıcaklık (T), Yağış (P), 1 gün ötelemeli Akış (Q_{t-1}), SVM RBF 2, SVM PK 2 ve MP5 2 modellerinde girdi olarak Ortalama sıcaklık (T), Yağış (P), 1 gün ötelemeli Yağış (P_{t-1}) 1 gün ötelemeli Akış (Q_{t-1}), SVM RBF3, SVM PK 3 ve MP5 3 modellerinde Ortalama sıcaklık (T), Yapış (P), 1 gün ötelemeli Yağış (P_{t-1}) 1 gün ötelemeli Akış (Q_{t-1}), ve 2 gün ötelemeli akış (Q_{t-2}), kullanılarak akış tahmini yapıldı. Oluşturulan

modeller arasında en iyi sonucu 3 numaralı modeller vermiştir. Modellerin uyumlulukları ve istatistiksel olarak hata değerleri Tablo 1'de gösterildiği gibidir

Tablo 1. Modellere göre sonuçların uyumluluk ve istatistiksel hata değerleri

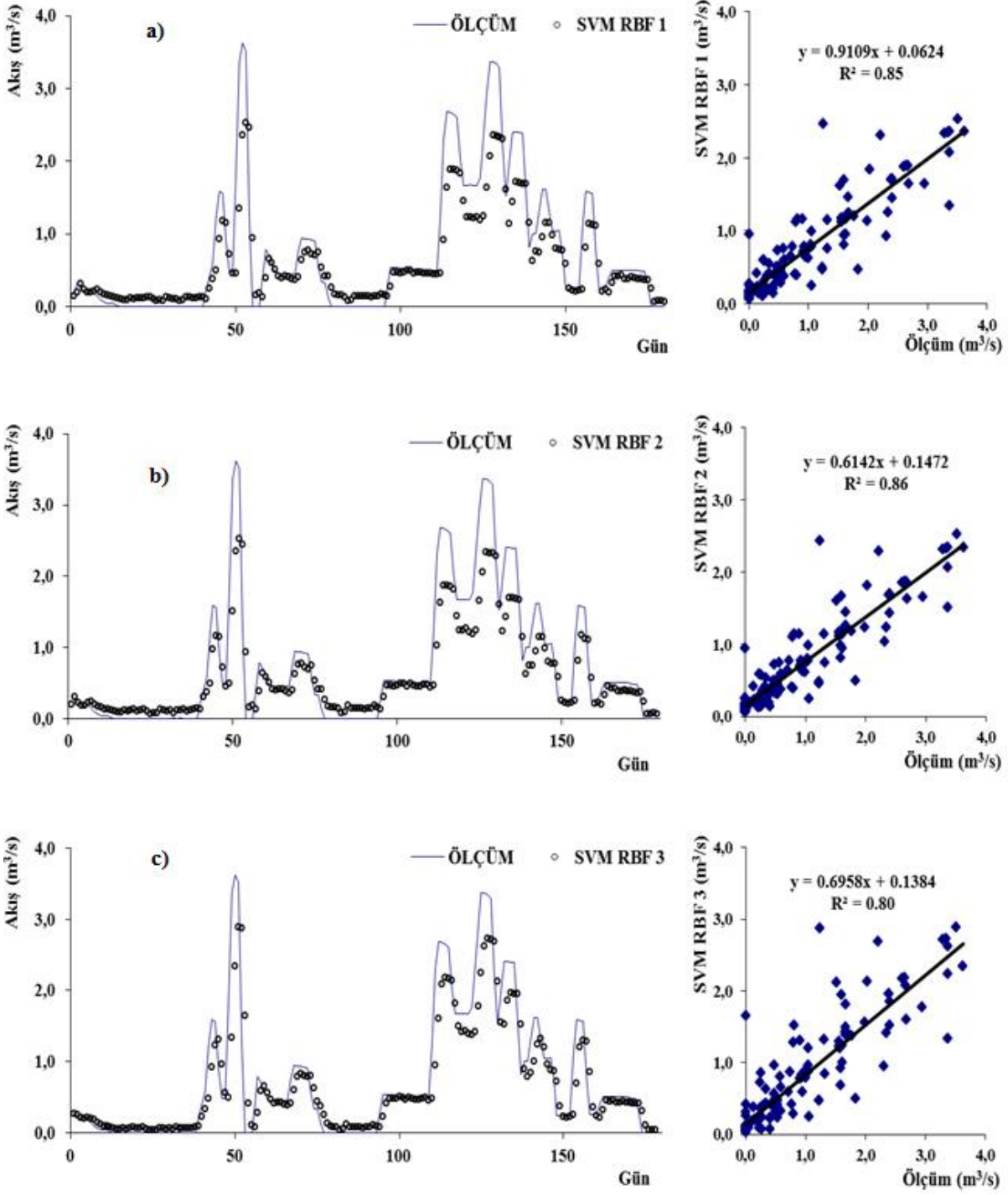
MODEL	GİRDİ	OMH	OKH	R ²
SVM RBF 1	P, T, Q_{t-1}	0.2754	0.1932	0.85
SVM RBF 2	P, P_{t-1}, T, Q_{t-1}	0.2747	0.1891	0.86
SVM RBF 3	$P, P_{t-1}, T, Q_{t-1}, Q_{t-2}$	0.253	0.183	0.80
SVM PK 1	P, T, Q_{t-1}	0.156	0.122	0.85
SVM PK 2	P, P_{t-1}, T, Q_{t-1}	0.158	0.120	0.86
SVM PK 3	$P, P_{t-1}, T, Q_{t-1}, Q_{t-2}$	0.142	0.090	0.89
M5T 1	P, T, Q_{t-1}	0.190	0.112	0.86
M5T 2	P, P_{t-1}, T, Q_{t-1}	0.186	0.099	0.88
M5T 3	$P, P_{t-1}, T, Q_{t-1}, Q_{t-2}$	0.175	0.072	0.91

Tüm modellerin performans değerlendirmesinde determinasyon katsayısı (R²), ortalama mutlak hata (OMH) ve ortalama karesel hataya (OKH) göre yapılmıştır. Test verilerinden elde edilen OMH, OKH ve R²'nin karşılaştırma parametreleri Tablo 1'de gösterilmektedir. Tablodaki sonuçlar, modellerin tahmin performansını ve gözlem verileri ile karşılaştırmak için kullanılmıştır. Oluşturulan üç modelin gerçek veriler ile olan ilişkilerini gösteren dağılım ve saçılım grafikleri SVM RBF için Şekil 3'de, SVM PK için ise Şekil 4'de, M5T için de Şekil 5'de gösterilmiştir.

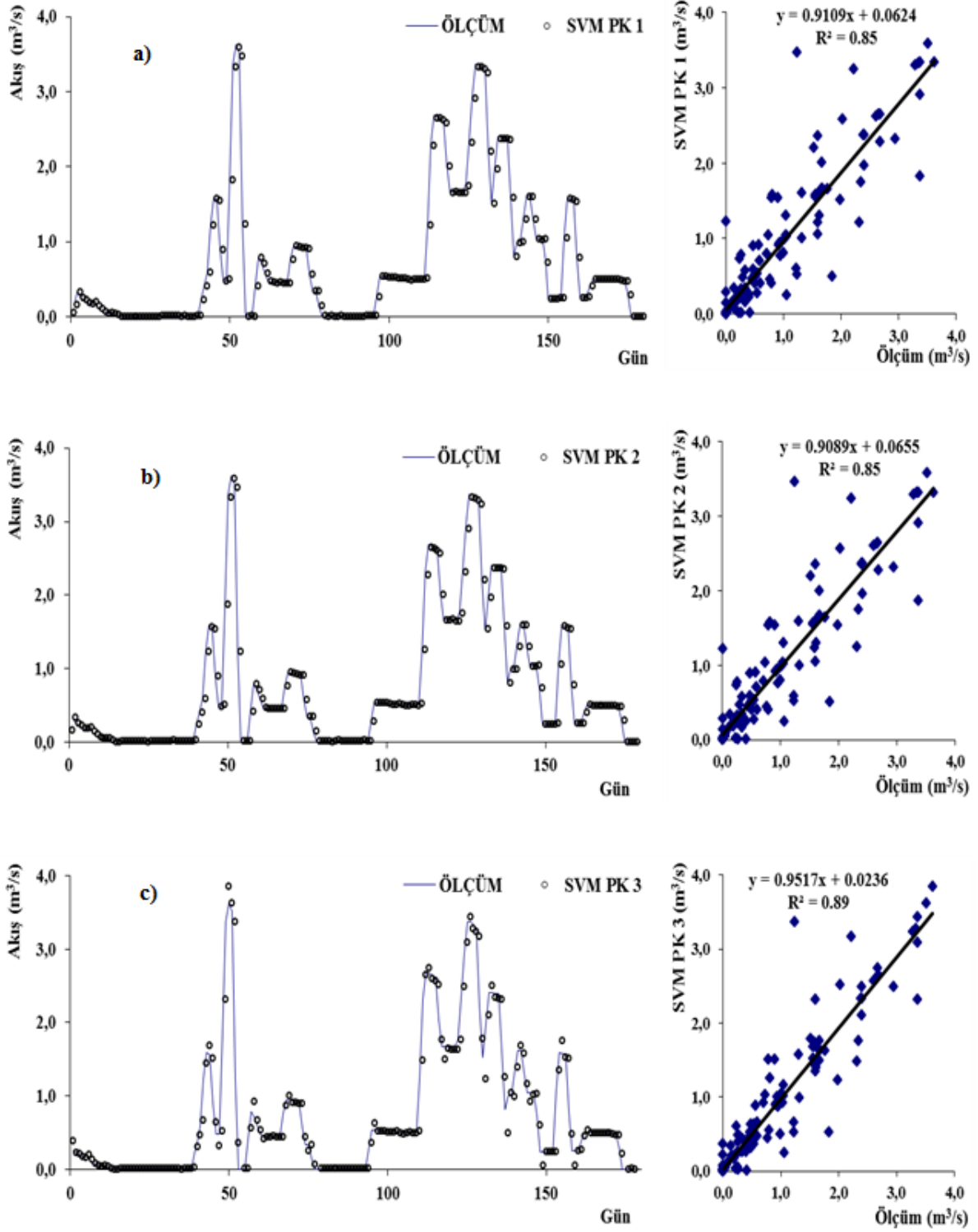
Şekil 3'de, SVM RBF modellerine bakıldığında korelasyon katsayısının yüksek olduğu ve yağış-akış tahminlerinin gerçek değerlere yakın olduğu görülmektedir. SVM RBF3 modeli, Tablo 1'de gösterilen edilen OMH, OKH ve R² (0.253, 0.183, 0.80) kriterlerine göre düşük hata, yüksek korelasyon göstermiş ve diğer SVM RBF modellerine göre daha iyi performans gösterdiği görülmüştür.

SVM PK modelleri için, dağılım ve saçılma grafikleri, Şekil 4'de gösterilmiştir. Test verileri için dağılım grafikleri analiz edildiğinde, SVM PK model sonuçları SVM RBF modellerine göre daha iyi bir performans gösterdiği görülmüştür. SVM PK3 modeli tahmini değerleri de gerçek değerlere yakındır. Determinasyon katsayısı en yüksek SVM PK 3 modelinde R² = 0.89 olarak elde edildi. SVM PK 3 model sonucu OMH, OKH ve R² (0.142, 0.090, 0.89) kriterlerine göre diğer SVM PK ve SVM RBF modellerine kıyasla daha iyi performans göstermiştir.

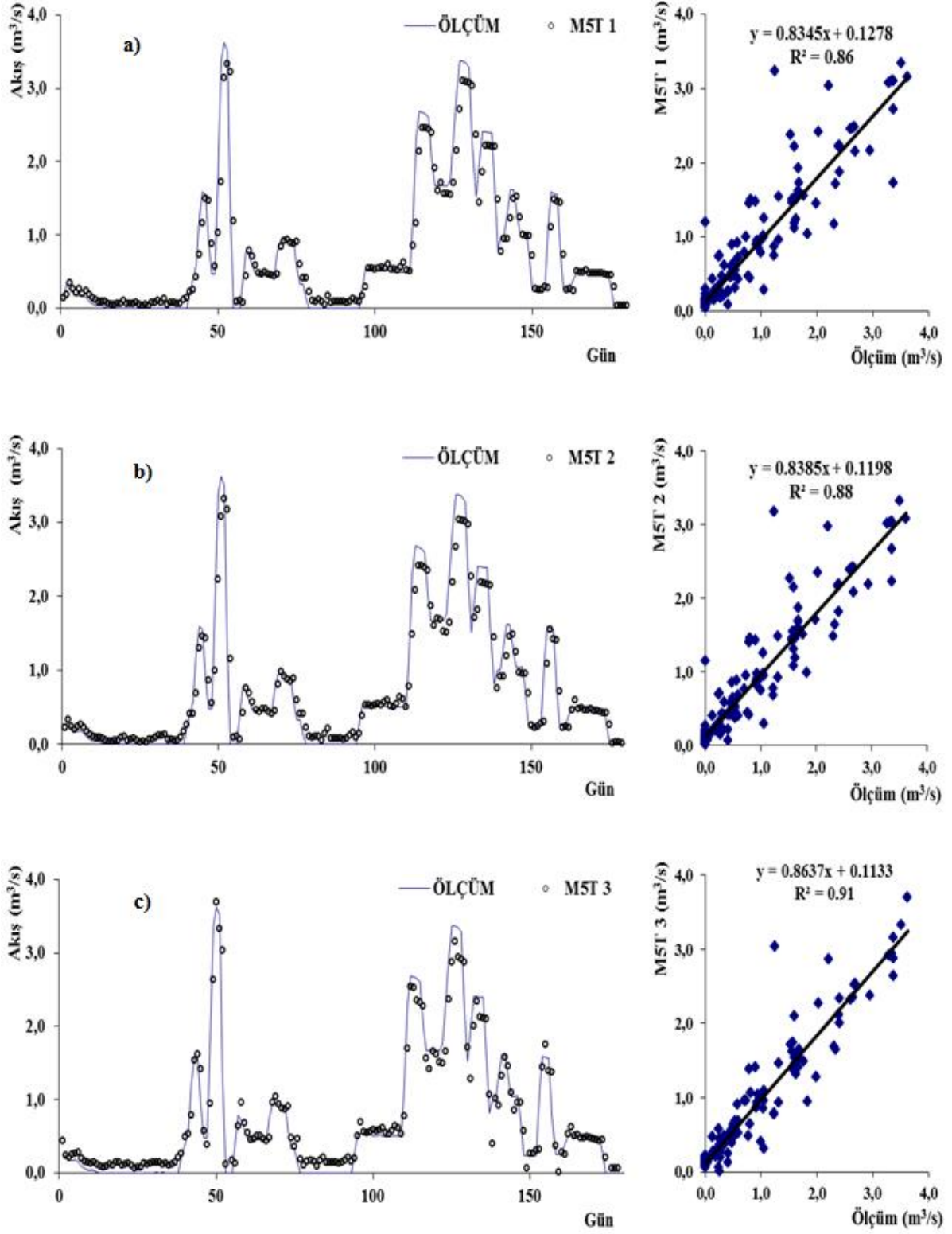
M5T modelleri için 731 günlük veriler değerlendirilmiş ve sonuçlar aşağıdaki gibi tanımlanmıştır. M5T modellerinin dağılım ve dağılım grafikleri, Şekil 5'de gösterilmektedir. Determinasyon katsayısı R² = 0.91 olarak elde edildi. Dağılım ve saçılım şekilleri incelendiğinde, tahmini değerler gerçek değerlere yakındır ve Tablo 1'de gösterilen OMH, OKH ve R² (0.175-0.072-0.91) kriter değerlerine baktığımızda, M5T 3 model sonucu, diğer M5T, SVM RBF ve SVM PK modellerine göre daha iyi sonuç göstermiştir.



Şekil 3. SVM RBF modelleri için dağılım ve saçılım diyagramları a) SVM RBF 1 b) SVM RBF 2 c) SVM RBF 3



Şekil 4. SVM PK modelleri için dağılım ve saçılım diyagramları a) SVM PK 1 b) SVM PK 2 c) SVM PK 3



Şekil 5. MST modelleri için dağılım ve saçılım diyagramları a) MST 1 b)MST 2 c) MST 3

Sonuçlar

Bu çalışmada, Stony Brook Rezervuarı havzasında bulunan 01104480 numaralı gözlem istasyonuna ait 731 günlük yağış, akış ve sıcaklık verileri Radyal Temelli Çekirdek Fonksiyonlu Destek Vektör Makineleri(SVM-RBF), Poli Çekirdek Fonksiyonlu Destek Vektör Makineleri (SVM-PK) ve M5 Karar Ağacı (M5T) metotlarında girdi olarak girilip yeni akış değerleri tahmin edilmiştir. Elde edilen sonuçlar ile gerçek sonuçlar karşılaştırılarak modeller arasındaki ilişki incelenmiştir. Modelleme sırasında 731 adet verinin %75'i eğitim aşamasında, kalan %25'i ise test aşamasında kullanılmıştır.

Yapılan çalışmalarda Poli Çekirdekli Fonksiyonlu Destek Vektör Makineleri (SVM-PK) yönteminin, Radyal Temel Fonksiyonlu Destek Vektör Makineleri (SVM-RBF) yöntemine göre daha az hatayla gerçeğe daha yakın sonuçlar ortaya koyduğu görülmüştür.

Tüm modelleri karşılaştırdığımızda M5 Karar Ağacı (M5T) modellerinin yağış-akış ilişkisi tahmininde SVM modellerine daha iyi performansa sahip olduğu görülmüştür. Düşük miktardaki hata (OMH, OKH) oranları ve yüksek determinasyon katsayısına (R^2) göre M5T 3 modelinin en iyi performans sağladığı görülmüştür.

Kaynaklar

Asefa, T., Kemblowski, M., Lall, U., Urroz, G. (2005). Support vector machines for nonlinear state space reconstruction: Application to the Great Salt Lake time series. *Water resources research*, 41(12).

Bhattacharya, B., Solomatine, D. P. (2005). Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing*, 63, 381-396.

Bray, M., Han, D. (2004). Identification of support vector machines for runoff modelling. *Journal of Hydroinformatics*, 6(4), 265-280.

Chen, H., Guo, J., Xiong, W., Guo, S., Xu, C. Y. (2010). Downscaling GCMs using the Smooth Support Vector Machine method to predict daily precipitation in the Hanjiang Basin. *Advances in Atmospheric Sciences*, 27(2), 274-284.

Cortes, C., Vapnik, V. (1995). Machine learning. *Support vector networks*, 20, 273-297.

Demirci M., Baltacı A. (2013), Prediction of suspended sediment in river using fuzzy logic and multilinear regression approaches. *Neural Computing and Applications*, 23(1), 145-151.

Demirci M., Unes F., Saydemir S. (2015a), Suspended sediment estimation using an artificial intelligence approach. *Sediment Matters*, Springer, 83-95.

Demirci M., Unes F., Aköz M.S. (2015b), Prediction of cross-shore sandbar volumes using neural network approach. *Journal of Marine Science and Technology*, 20(1), 171-179.

Demirci M., Unes F., Aköz M.S. (2016), Determination of nearshore sandbar crest depth using neural network approach. *International Journal of Advanced Engineering Research and Science*, 3(12), 133-140.

Demirci M., Tasar B., Kaya Y.Z., (2018), Estimation of Groundwater Level Fluctuations Using Neuro-Fuzzy and Support Vector Regression Models. *Int. J. Adv. Eng. Res. Sci.* 5, 206–211. doi:10.22161/ijaers.5.12.29

Demirci, M., Unes, F., Kaya, Y. Z., Tasar, B., Varcin, H. (2018). Modeling of dam reservoir volume using adaptive neuro fuzzy method. *Aerul si Apa. Componente ale Mediului*, 145-152.

Dibike, Y. B., Velickov, S., Solomatine, D., Abbott, M. B. (2001). Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15(3), 208-216.

Ergezer, H., Dikmen, M., Özdemir, E. (2003). Yapay sinir ağları ve tanıma sistemleri. *PiVOLKA*, 2(6), 14-17.

Fernando, D. A. K., Jayawardena, A. W. (1998). Runoff forecasting using RBF networks with OLS algorithm. *Journal of hydrologic engineering*, 3(3), 203-209.

Hamidi, N. ve Demirci, M., (2006). "Yapay Sinir Ağları ve Regresyon Teknikleri ile Botan Çayı Aylık Taşınan Sediment Miktarının Modellenmesi". *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, Cilt.21, No.1–2 (Haziran-Aralık), 193–207, Adana.

Hamidi, N. and Kayaalp, N., (2008). "Estimation of the Amount of Suspended Sediment in the Tigris River using Artificial Neural Networks". *CLEAN-Soil, Air, Water*, 36(4), 380–386.

Hosseini, S. M., Mahjouri, N. (2016). Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling. *Applied Soft Computing*, 38, 329-345.

Hsu, C. W., Chang, C. C. Lin, C. J. (2003) A practical guide to support vector classification. Tech. Report, Dept Computer Sci. & Info.

- Engng, National Taiwan University, Taiwan, China.
- Hsu, K. L., Gupta, H. V., Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, 31(10), 2517-2530.
- Kaya, Y. Z., Mamak, M., Unes, F. (2016). Evapotranspiration Prediction Using M5T Data Mining Method. *International Journal of Advanced Engineering Research and Science*, 3(12).
- Kaya, Y. Z., Üneş, F., Demirci, M., Taşar, B., Varçin, H. (2018). Groundwater Level Prediction Using Artificial Neural Network and M5 Tree Models. *Aerul si Apa. Componente ale Mediului*, 195-201.
- Khan, M. S., Coulibaly, P. (2006). Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering*, 11(3), 199-205.
- Khalil, A. F., McKee, M., Kemblowski, M., Asefa, T., Bastidas, L. (2006). Multiobjective analysis of chaotic dynamic systems with sparse learning machines. *Advances in Water Resources*, 29(1), 72-88.
- Mason, J. C., Price, R. K., Tem'Me, A. (1996). A neural network model of rainfall-runoff using radial basis functions. *Journal of Hydraulic Research*, 34(4), 537-548.
- McCulloch, W. S., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- Minns, A. W., Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrological sciences journal*, 41(3), 399-417.
- Nieto, P. G., Torres, J. M., Fernández, M. A., Galán, C. O. (2012). Support vector machines and neural networks used to evaluate paper manufactured using Eucalyptus globulus. *Applied Mathematical Modelling*, 36(12), 6137-6145.
- Quinlan, J. R. (1992). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* 92, 343-348.
- Rao, M., Fan, G., Thomas, J., Cherian, G., Chudiwale, V., Awawdeh, M. (2007). A web-based GIS Decision Support System for managing and planning USDA's Conservation Reserve Program (CRP). *Environmental Modelling & Software*, 22(9), 1270-1280.
- Sattari, M. T., Pal, M., Apaydin, H., Ozturk, F. (2013). M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey. *Water Resources*, 40(3), 233-242.
- Solomatine, D. P., Xue, Y. (2004). M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. *Journal of Hydrologic Engineering*, 9(6), 491-501.
- Tasar B., Unes F., Demirci M., Kaya Y.Z. (2018), Yapay sinir ağları yöntemi kullanılarak buharlaşma miktarı tahmini. *DÜMF Mühendislik Dergisi*, 9(1), 543-551.
- Tasar B., Kaya Y. Z., Varçin H., Unes F., Demirci M. (2017), Forecasting of suspended sediment in rivers using artificial neural networks approach. *International Journal of Advanced Engineering Research and Science*, 4(12), 79-84.
- Tripathi, S., Srinivas, V. V., Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology*, 330(3-4), 621-640.
- Turhan, E., 2012. Seyhan Havzası'nın Yağış- Akış İlişkisinin Yapay Sinir Ağları Yöntemi ile Modellenmesi, Yüksek Lisans Tezi, Adana.
- Unes F. (2010a), Dam reservoir level modeling by neural network approach: a case study. *Neural Network World*, 4(10), 461.
- Unes F. (2010b), Prediction of density flow plunging depth in dam reservoirs: an artificial neural network approach. *Clean-Soil, Air, Water*, 38(3), 296-308.
- Unes F., Demirci M. (2015), Generalized Regression Neural Networks For Reservoir Level Modeling. *International Journal of Advanced Computational Engineering and Networking*, 3, 81- 84.
- Unes F., Yildirim S., Cigizoglu H.K., Coskun H. (2013), Estimation of dam reservoir volume fluctuations using artificial neural network and support vector regression. *Journal of Engineering Research*, 1(3), 53-74.
- Unes F., Demirci M., Kişi Ö. (2015), Prediction of millers ferry dam reservoir level in USA using artificial neural network. *Periodica Polytechnica Civil Engineering*, 59(3), 309-318.
- Unes F., Gumuscan F.G., Demirci M. (2017), Prediction of Dam Reservoir Volume Fluctuations Using Adaptive Neuro Fuzzy Approach. *EJENS*, 2(1), 144-148.
- Üneş, F., Demirci, M., İspir, E., Kaya, Y. Z., Mamak, M., Tasar, B. (2017). Estimation of Groundwater Level Using Artificial Neural Networks: a Case Study of Hatay-Turkey. In *Environmental Engineering. Proceedings of the International Conference on Environmental Engineering. ICEE* (Vol. 10, pp. 1-6). Vilnius

Gediminas Technical University, Department of Construction Economics & Property.

- Uneş F., Demirci M., Mertcan Z., Taşar B., Varçin H., Ziya Y. (2018a). Determination of Groundwater Level Fluctuations by Artificial Neural Networks . *Natural and Engineering Sciences*, 3(3), Supplement, 35-42.
- Uneş F., Doğan S., Taşar B., Kaya Y., Demirci M. (2018b), The Evaluation and Comparison of Daily Reference Evapotranspiration with ANN and Empirical Methods. *Natural and Engineering Sciences*, 3(3), Supplement, 54-64.
- Unes F., Bölük O., Kaya Y. Z., Tasar B., Varçin H. (2018c), Estimation of Rainfall-Runoff Relationship Using Artificial Neural Network Models for Muskegon Basin. *Journal of Engineering Research*.
- USGS.gov | Science for a changing world [WWW Document], n.d. URL <https://www.usgs.gov/>
- Vapnik, V. The Nature of Statistical Learning Theory; Springer: New York, NY, USA, 1995.
- Vapnik, V., Golowich, S., Smola, A. (1997) Support vector method for function approximation, regression estimation, and signal processing. *In Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, pp. 281–287.
- Witten, I.H., Frank, E. (2005). Data mining: Practical machine learning tools and techniques with java implementations, Morgan Kaufmann, San Francisco, CA.

Estimation of rainfall-runoff relationship using support vector machines and M5 decision tree methods

Extended abstract

In terms of efficient use of water resources and prevention of disasters, rainfall runoff relation is very important.. In this study, the data obtained from USA Waltham Massachusetts Stony Brook Reservoir basin was used. Rainfall -Runoff relation was predicted with using Support Vector Machine (SVM) and M5 Decision Trees (M5T) Method.

The data were obtained from the Stony Brook River, which supplies the Stony Brook reservoir at 42 ° 21'20 "Latitude and 71 ° 15'56" Longitude near the Waltham Dam in the US state of Massachusetts in the United States. The Stony Brook basin has a drainage area of 61.8 km² and a storage capacity of 1,340,000 m³. Station data from 01104480 taken from the United States Geological Research Institute (USGS) contains 731 days of precipitation, flow and temperature information between 07.05.2011 and 07.05.2013.

In this study 731 data was collected to predict the rainfall-runoff relationship from the station. 550 of these data were used in the training phase, remaining 181 data were used in the test phase. Predictions were made by using Support Vector Regression with radial basis functions (SVM-RBF), Support Vector Regression with poly kernel (SVM-PK) and Decision Tree Methods (M5T). Three evaluations were made in all three models. Estimation results of the models were compared according to R² (Determination coefficient), Mean Absolute Error (MAE) and Mean Square Error (MSE) criteria.

Average temperature (T), Precipitation (P), 1 day shift Flow (Q_{t-1}) were taken as input in SVM RBF 1, SVM PK 1 and MP5 1 models, average temperature (T), precipitation (P), lagged-1 day precipitation (P_{t-1}), lagged-1 day flow (Q_{t-1}) were taken as input in SVM RBF 2, SVM PK 2 and MP5 2 models, average temperature (T), precipitation (P), lagged-1

day precipitation (P_{t-1}) lagged-1 days flow (Q_{t-1}) and lagged-2 day flow (Q_{t-2}) were taken as input in SVM RBF 3, SVM PK 3 and MP5 3 models. The best results were the models of model 3.

The results obtained were compared with the actual results. As a result of the study, it has been seen that Support Vector Regression with poly kernel (SVM-PK) method has less realistic results with less error than Support Vector Regression with radial basis functions (SVM-RBF)

It is seen that M5 Decision Tree (M5T) models have better performance in the predicting rainfall-runoff relationship than the all SVM models. The M5T 3 model provided the best performance according to the low amount of error (MAE, MSE) ratios and high determination coefficient (R²).

Keywords: *Rainfall - Runoff, Prediction, Support Vector Machine, M5 Decision Tree*