

Book of Full Texts

Afet Yönetimi/Disaster Management

IESKO 2019
VI. International Earthquake Symposium Kocaeli 2019
September 25-27, 2019, Kocaeli, Turkey

Kantil ve Lojistik Regresyona Dayali Heyelan Duyarlilik Haritalama Yöntemleri için Özellik Altkümesi Seçiminin Analizi

Feature Subset Analysis for Landslide Susceptibility Mapping Based On Quantile and Logistics Regression Methods

Emrehan Kutluğ Şahin^{1*+}

¹İnşaat Mühendisliği/Mühendislik Fakültesi, Bolu Abant İzzet Baysal Üniversitesi, Bolu, Türkiye

*Sorumlu yazar: emrehansahin@ibu.edu.tr

+Konuşmacı: emrehansahin@ibu.edu.tr

Sunum Türü: Sözlü

Sunum Dili : Türkçe

Özet

İnsanlar tarafından yapılan altyapılara zarar veren ve birçok yönden insan hayatını olumsuz bir biçimde etkileyen heyelanlar tarihte en yıkıcı doğal afetlerden birisi olarak nitelendirilmektedir. Afet yönetimindeki büyük önemi nedeniyle doğruluğu yüksek ve güncel heyelan duyarlılık haritalarının üretilmesi küresel ve yerel ölçekli çalışmalar için esas teşkil etmektedir. Literatürde, heyelan duyarlılık haritalarının üretiminde çeşitli stratejiler, algoritmalar, yöntemler ve bunların farklı kombinasyonların kullanımı önerilmektedir. Buna karşın heyelan duyarlılığının değerlendirilmesinde birtakım sınırlamalar ve belirsizlikler mevcuttur. Bu çalışmada, Sinop ilinin Ayancık ilçesine ait heyelan duyarlılık haritasının üretilmesi için Adımsal Regresyon (AR) metodu yardımıyla optimum faktör modeli tespit edilmiştir. AR yöntemiyle belirlenen modelden sonra heyelan duyarlılık haritalamasında kantitatif ve lojistik regresyon yöntemleri kullanılmıştır. Çalışma bölgesi için 12 faktör (bakı, akarsuya olan uzaklık, yükseklik, litoloji, arazi örtüsü/kullanımı, NDVI, plan eğrisi, profil eğrisi, sediment taşıma indeksi, akarsu güç indeksi, eğim ve topoğrafik nemlilik indeksi) belirlenmiştir. Bu faktörler içerisinde AR metodu yardımıyla 8 optimum faktör (arazi örtüsü/kullanımı, NDVI, yükseklik, bakı, litoloji, eğim, topoğrafik nemlilik indeksi ve sediment taşıma indeksi) tespit edilmiştir. Genel doğruluk değerleri incelendiğinde kantil regresyon metodu Model-8 için %77,11 ve tüm faktörleri içeren Model-12 için %76,99 ve lojistik regresyon yöntemi ise Model-8 için %77,49 ile tüm veri seti için %77,93 oranında doğruluk değerleri vermiştir. Üretilen duyarlılık haritaları içerisinde en iyi performansı değerini Model-12 ve lojistik regresyon algoritması ile üretilmiş heyelan duyarlılık haritasını olduğu görülmüştür. Fakat, Model-8 ve Model-12 arasındaki istatistiksel anlamlılık farklılıkları McNemar'ın testi ile incelendiğinde, bu iki model arasındaki farklılıkların önemsiz olduğu bulunmuştur. Sonuç olarak, SR tarafından elde edilen optimum faktör modeli olan Model-8, bu çalışma alanı için lojistik regresyon süreci ile en uygun heyelan duyarlılık haritası olarak bulunmuştur.

Anahtar kelimeler: Kantil regresyon, adımsal regresyon, lojistik regresyon, heyelan duyarlılık haritası.

Abstract

Landslides have been qualified as a one of the most destructive natural disasters in history that cause damage to man-made structures and negatively affect human life in many ways. Due to its great importance for hazard management, producing accurate and up-to-date landslide susceptibility maps is essential for many local- to global-scale studies. For this purpose, various strategies, algorithms, methods and their different combinations have been suggested in the literature. However, there are several limitations and uncertainties in assessing landslide susceptibility. In this study, the optimum factor model was determined by using the Stepwise Regression (SR) method to produce the landslide susceptibility map of Ayancık district of Sinop province. Quantitative and logistic regression methods were utilized in the production of the landslide susceptibility mapping after the model determined by the SR method. The twelve factors were determined for the study area (aspect, distance from river, elevation, lithology, land cover / land use, NDVI, plan curvature, profile curvature, sediment transport index, stream power index, slope and topographic wetness index). 8 optimum factors (land cover / land use, NDVI, elevation, aspect, lithology, slope, topographic wetness index and sediment transport index) have been determined by SR method among these main factor clusters. When the overall accuracy results were analyzed, the quantile regression method was 77.11% for Model-8 and 76.99% for Model-12 with all the factors and the logistic regression method also was found 77.99% accuracy for Model-8 and 77.93% for the entire data set. When the best performance results were analyzed for the produced landslide susceptibility maps, it was found that Model-12 based on logistic regression algorithm was selected the best model for this study. However, when statistical significance differences between Model-8 and Model-12 was analyzed by McNemar's test, it was found that the differences between these two models were insignificant. As a result, Model-8, which is the optimum factor model obtained by SR, has been found to be the most suitable landslide susceptibility map for this study area by logistic regression process.

Keywords: Quantile regression, stepwise regression, logistic regression, landslide susceptibility map.

1. Giriş

Heyelan duyarlılık haritalarının üretilmesindeki en önemli işlem adımlarından biri heyelana sebep olabilecek eğim, bakı, yükseklik ve benzeri faktörlerin belirlenmesi ve bu faktörlerin kullanıldığı optimum harita üretim metodunu belirlenmesidir. Bu kapsamda doğal etmenlere dayalı faktörler ile birlikte insan kaynaklı birçok faktörün bir arada değerlendirilmesi gerekmektedir. Ayrıca duyarlılık haritası üretiminde optimum faktör kümesinin belirlenmesi ile birlikte modellemenin oluşturulabilmesi için en uygun harita üretim metoduna da karar verilmelidir. Heyelan duyarlılık haritası üretimi, çok sayıda faktörün, farklı harita üretim metodları ile birlikte bir bütün halinde değerlendirilmesini gerektiren önemli bir çalışma konusudur. Söz konusu süreç içerisinde ayrıca karar vericilerin sezgisel, deneme-yanılma veya benzer çalışmalarda neler tercih ediliyor gibi yaklaşımlara başvurmaları ile yanlış kararların verilmesi riski değerlendirme sürecini etkileyen önemli bir etmendir. Ayrıca, heyelan duyarlılık haritalarında kullanılan faktörlerin seçimi ve hangi harita üretimi metodunun kullanılması konusunda henüz evrensel bir uzlaşma bulunmadığı söylenebilir (Gokceoglu ve Ercanoğlu, 2001; Sakellariou ve Ferentinou, 2001).

Heyelan alanlarının tespitinde kullanılabilecek doğru ve güncel bir heyelan duyarlılık haritasının afet planlamalarında önemi büyüktür. Özellikle gelecekte meydana gelecek heyelan afetleri için gerekli tedbirlerin alınmasında karar vericiler ve ilgili kurumların planlama süreçleri için büyük önem taşımaktadır. Bu çalışmada üretilecek heyelan duyarlılık haritasında 12 farklı faktör dikkate alınmıştır. Söz konusu faktör kümesi içinde heyelan duyarlılık haritası üretiminde

kullanılmak üzere optimum model tespiti için adimsal regresyon (AR) metodundan kullanılmıştır. Optimum faktör modelinin elde edilmesinden sonra heyelan duyarlılık haritası üretiminde kantil regresyon (KR) ve lojistik regresyon (LR) metodundan faydalanılmıştır. Üretilen heyelan duyarlılık haritalarının performanslarının değerlendirilmesi amacıyla genel doğruluk ve AUC değerleri karşılaştırılmıştır. Ayrıca çalışmada optimum model ve diğer veri seti ile üretilmiş haritaların birbirleriyle olan farklılıkları ve istatistiksel anlamlılıklarının irdelenmesi için McNemar's testi kullanılmıştır.

2. Çalışma Alanı

Batı Karadeniz bölgesi Sinop ili içerisinde bulunan Ayancık ilçesi çalışma bölgesi olarak tespit edilmiştir. Ayancık ilçesinin kuzeyinden Karadeniz, Bartın ili, batısında Türkeli ilçesi, doğusunda Erfelek ilçesi ve güneyinde Boyabat ilçesi ve Kastamonu ili bulunmaktadır. Çalışma alanı yaklaşık 886 km² bir alanı kapsamaktadır. Yıllık ortalama yağış oranı 676 kg/m² ve çalışma alanının büyük bir kısmını eğimli araziler teşkil etmektedir. Çalışma bölgesinin meteorolojik olarak yoğun yağış alması, eğimli arazi yapısı ve heyelana duyarlı toprak yapısı nedeniyle toprak kayması, heyelan ve benzeri doğal afetlere karşı meyilli bir bölge konumundadır. Çalışma alanındaki yükseklik 0 m ile 1.670 m, arazi eğimi 0° ile 72° arasında değişmektedir.

3. Yöntemler

3.1. Adimsal Regresyon

Çok sayıda verinin değerlendirileceği karmaşık ve kompleks durumlarda veri seti içerisindeki ilgisiz verilerin elenmesi işlemi regresyona dayalı olarak hesaplayan metoda adimsal regresyon denilmektedir (Neter vd., 1996; Saltelli vd., 2000). Regresyon analizlerinin genel işleyişi iki veya daha fazla bağımsız değişken ile bir bağımlı değişken arasındaki ilişkiyi belirleyerek ve söz konusu bu ilişkiyi bir modelle tanımlayabilmektir. Adimsal regresyon analizlerinde ise bağımsız değişken olarak tanımlanan faktör modelinin içerisindeki en ideal veya bağımlı değişken ile arasında en yüksek ilişkiyi barındıran faktörün belirlenmesinde bağımsız değişkenlerin kendi aralarındaki korelasyonun düşük ama bağımlı değişken ile korelasyonunun yüksek olması şartı aranmaktadır (Şahin, 2018).

3.2. Lojistik Regresyon

Yer bilimci araştırmalarında olduğu gibi heyelan duyarlılık haritalarının üretilmesinde de en çok kullanılan istatistiksel analiz yaklaşımlarından biri lojistik regresyon (LR) yöntemidir. LR yönteminin temel amacı, bağımlı bir değişken ile bağımsız birçok değişken arasında çok değişkenli bir regresyon ilişkisi kurmayı sağlamaktır (Lee, 2005). Bu regresyon modelinde bağımlı değişken, heyelan olayının olması ya da olmaması durumunu gösterirken, bağımsız değişkenler ise modelde kullanılan ve heyelan olayına etki eden faktörlerdir. Heyelana duyarlı alanların tespiti gibi çalışmalarda bağımlı değişken girdi verisi olmayan alanlar için 0 ve heyelan olan alanlar için ise 1 olmalıdır.

3.3 Kantil Regresyon

Kantil regresyon temel fikri ilk olarak Koenker ve Bassett (1978) tarafından önerilmiştir. Kantil regresyon temel olarak, koşullu kantillerin bağımsız değişkenlerin bir fonksiyonu olarak modellenmesidir. Kantil regresyon metodunu geleneksel regresyon modellerinde ayıran en büyük özelliği bağımlı değişkenin koşullu ortalamasındaki değişimleri açıklamaya çalışmak yerine kantil regresyon koşullu ile kantillerdeki değişimleri açıklamaktadır. Bu yönüyle geleneksel regresyona göre daha esnek ve araştırmanın niteliğine göre farklı kantiller kullanılarak optimum çözüm sağlanabilir (Akar, 2013). Literatürde kantil katsayılarının $\tau = 0.25, 0.50, 0.75, 0.95$ olarak tanımlanmıştır ve bu katsayılar bağımlı değişkenin dağılımı boyunca

farklı etkiler altında olduğu için sabit etki panel katsayılarından anlamlı şekilde farklı olabileceğine dikkat çekilmektedir (Uğur ve Özocaklı, 2018).

4. Uygulama

Çalışma alanında toplam 165 adet heyelan poligonu sayısallaştırılarak CBS ortamına aktarılmıştır. Poligon türündeki heyelan verileri 30x30 m'lik raster veri formatına dönüştürülmüştür. Heyelan envanteri içerisindeki heyelan olan alanlar için "1" ve heyelan olmayan alanlar içinse "0" değerleri atanmıştır. Çalışmada, optimum faktör seçimi, heyelan duyarlılık haritası üretimi ve doğrulama işlemlerinde %70 eğitim ve %30 test verisi seti oluşturulmuştur. Duyarlılık haritalarının yeniden sınıflandırılması işleminde eşit dağılım (quantile) esasına dayanılarak 5 sınıfta (çok yüksek, yüksek, orta, düşük ve çok düşük) bölümlendirilmiştir. Sonuç haritaların birbirleriyle olan farklarının istatistiksel tespiti ve optimum haritanın karar verilmesi aşamasında McNemar's istatistik anlamlılık testi kullanılmıştır.

4.1. Adımsal regresyon ile optimum faktör model tespiti

Tüm faktör seti içerisinde birbirleriyle yüksek korelasyonu eleyen ve eğitim seti ile tam uyumlu modellerin tespiti ve faktör modelinin elde edilmesi amacıyla SPSS v.20 yazılımı kullanılarak AR işlemi yapılmıştır. (Tablo 1). Optimum model analizde, bağımlı değişken üzerinde katkısı az olan drenaj yoğunluğu, akarsuya uzaklık, plan ve profil eğrisi ve akarsu güç indeksi faktörleri optimum model dışında kalmıştır. Sonuç olarak 12 faktör setinden oluşan veri seti içerisinde 4 faktör elenerek 8 adet faktörden oluşan Model-8 optimum faktör setine sahip model olarak kabul edilmiştir.

Tablo 1. Adımsal regresyon analizi sonucu elde edilen modeller.

Model	R	R ²	Düzeltilmiş R ²	Standart Tahmin Hatası
1	0,414 ^a	0,171	0,170	0,454
2	0,469 ^b	0,220	0,218	0,441
3	0,525 ^c	0,276	0,273	0,425
4	0,555 ^d	0,308	0,304	0,416
5	0,566 ^e	0,321	0,317	0,412
6	0,576 ^f	0,331	0,327	0,409
7	0,583 ^g	0,340	0,334	0,407
8	0,588 ^h	0,346	0,340	0,405
Model-1: AÖAK			Model-5: AÖAK, NDVI, yükseklik, bakı, litoloji	
Model-2: AÖAK, NDVI			Model-6: AÖAK, NDVI, yükseklik, bakı, litoloji, eğim	
Model-3: AÖAK, NDVI, yükseklik			Model-7: AÖAK, NDVI, yükseklik, bakı, litoloji, eğim, TII	
Model-4: AÖAK, NDVI, yükseklik, bakı			Model-8: AÖAK, NDVI, yükseklik, bakı, litoloji, eğim, TII, STI	

4.2. Heyelan duyarlılık haritası üretimi

Çalışmada AR analizin vermiş olduğu optimum faktör setine sahip Model-8 kullanılarak KR ve LR metotları yardımıyla heyelan duyarlılık haritaları üretilmiştir. Üretilen heyelan duyarlılık haritalarının performanslarının analiz edilmesi amacıyla genel doğruluk ve AUC (Eğri Altında Kalan, Area Under Curve) değerlerine bakılmıştır. Optimum faktör modeli yaklaşımının heyelan duyarlılık haritası üretimindeki etkilerinin karşılaştırılması amacıyla AR yaklaşımıyla elde edilen Model-8 ve tüm faktörlerinde bulunduğu Model-12 için performans testleri yapılmıştır (Tablo 2).

Üretilen modellerin birbirleriyle olan ilişkilerini ve istatistiksel olan anlamlarının tespiti için McNemar's istatistik anlamlılık testi kullanılmıştır (Çizelge 3). Hesaplanan istatistik kritik değerinin %95 güven aralığı için 3,84'ten büyük olması üretilen tematik harita doğrulukları arasındaki farkın istatistiksel olarak anlamlı olduğu, değilse anlamsız olduğu söylenebilir. Test sonuçları incelendiğinde LR algoritmalarının her iki model türünde (Model-8 ve Model-12) aralarındaki farkın istatistiksel olarak anlamsız olduğu görülmüştür. Diğer bir taraftan KR

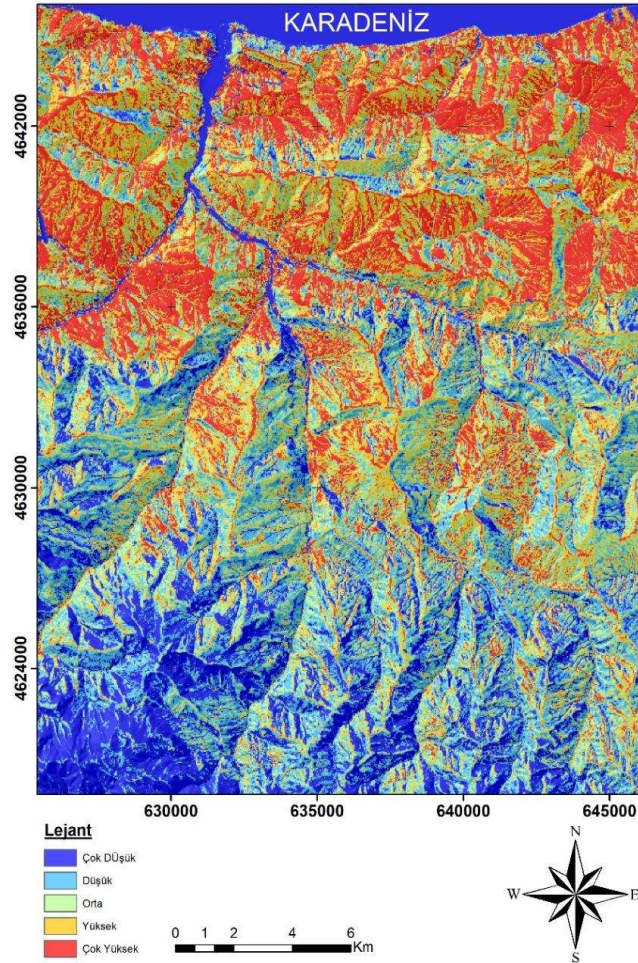
algoritmalarının her iki model türünde aralarındaki farkın istatistiksel olarak anlamlı olduğu görülmüştür. LR algoritması Model-8 ile üretilmiş duyarlılık haritasının optimum sonuçlar verdiği ve Model-12 ile aynı özelliklere sahip bir tematik harita olduğu tespit edilmiştir.

Tablo 2. Heyelan duyarlılık haritaları için elde edilen genel doğruluk ve AUC değerleri.

	Genel Doğruluk	AUC
KR Model-8	77,11	0.7695
KR Model-12	76,99	0.7689
LR Model-8	77,49	0.7737
LR Model-12	77,93	0.7779

Tablo 3. Duyarlılık modellerinin McNemar's testi kullanılarak istatistiksel karşılaştırılması.

	McNemar's (z=3,84)	Anlamlılık (%95)
LR ^{Model-8} / LR ^{Model-12}	3,082	Anlamsız
LR ^{Model-8} / KR ^{Model-8}	4,863	Anlamlı
LR ^{Model-8} / KR ^{Model-12}	0,289	Anlamsız
LR ^{Model-12} / KR ^{Model-8}	0,830	Anlamsız
LR ^{Model-12} / KR ^{Model-12}	3,486	Anlamsız
KR ^{Model-8} / KR ^{Model-12}	16,673	Anlamlı



Şekil 1. AR'nin belirlediği 8 faktörlü LR metoduna dayalı heyelan duyarlılık haritası

Adımsal regresyon seçimi ile elde edilen sıralama sonrası lojistik regresyon metodu kullanılarak en iyi doğruluğun elde edildiği heyelan duyarlılık haritası Şekil 1'de gösterilmiştir. Heyelan

duyarlılık haritası sonrası olası heyelan alanlarının risk dağılımı incelendiğinde çok yüksek ve yüksek risk değerlerinin çalışma alanının yoğunlukla kuzey kısımlarında baskın olduğu görülmektedir. Duyarlılık haritasının güney kısımlarının büyük bir bölümünde heyelan risk derecelendirmesinin yüksek ve orta yüksek olduğu görülmüştür. Son olarak yapılan analizler sonrası Ayancık ilçesinin güney kesimlerinin büyük bir kısmının düşük ve çok düşük heyelan duyarlılığına sahip olduğu tespit edilmiştir.

5. Sonuç

Bu çalışmada Sinop ili Ayancık ilçesi için üretilecek heyelan duyarlılık haritasında 12 faktörden oluşan veri seti dikkate alınmıştır. Mevcut faktör seti adimsal regresyon yaklaşımı ile tüm faktör seti içerisinde en uygun faktörlerden oluşan 8 faktörlü (arazi örtüsü/kullanımı, NDVI, yükseklik, bakı, litoloji, eğim, topoğrafik nemlilik indeksi ve sediment taşıma indeksi) optimum model belirlenmiştir. Optimum faktör modelinin belirlenmesinde sonra lojistik ve kantil regresyon analizleri yardımıyla heyelan duyarlılık haritaları üretilmiştir. Harita üretim metotları sonrası elde edilen haritalar içerisinde en iyi genel doğruluğu %77,49 ile 8 adet faktörden oluşan lojistik regresyon metoduna ait model vermiştir. Her ne kadar 12 faktör setiyle oluşan model daha iyi genel doğruluk ürettiği tespit edilmesine rağmen McNemar's test sonuçları ile de bu iki model arasındaki farkın istatistiksel olarak anlamlı olmadığı görülmüştür. Sonuç olarak bu çalışmada, çok sayıda faktör ile optimum faktör aramak ve otomatik seçim yaklaşımları ile optimum faktör setinin tespitinin mümkün olduğu ve sadece seçilen faktörler ile üretilen heyelan duyarlılık haritalarının genel doğruluk ve model performansının çok faktör seti ile üretilen modellerle eş veya yüksek olduğu istatistiksel anlamlılık testleri ile de ortaya konulmuştur.

Katkı Belirtme

Bu çalışma, TÜBİTAK tarafından 118Y090 numaralı proje ile desteklenmiştir.

Kaynaklar

- Akar, C., 2013. Gelişmekte Olan Piyasalarda Finansal Piyasa İstikrarının Kantil Regresyon Yöntemiyle Test Edilmesi. *Doğuş Üniversitesi Dergisi*, 14 (1), 1-9.
- Gökceoğlu, C., Ercanoğlu, M., 2001. Heyelan duyarlılık haritalarının hazırlanmasında kullanılan parametrelere ilişkin belirsizlikler. *Yerbilimleri*, 23, 189-206.
- Koenker, R., Bassett, G., 1978. Regression Quantiles. *Econometrica*, 46 (1), 33-50, doi: 10.2307/1913643.
- Lee, S., 2005. Application of Logistic Regression Model and Its Validation for Landslide Susceptibility Mapping Using GIS and Remote Sensing Data, *International Journal of RemoteSensing*, 26, 1477-1491, doi: 10.1080/01431160412331331012.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W., 1996. *Applied linear regression models*, Richard D. Irwin, Inc., Burr Ridge, Illinois.
- Şahin, E.K., 2018. Heyelan Duyarlılık Haritası İçin Adımsal Regresyona Dayalı Faktör Seçme Yönteminin etkinliğinin Araştırılması. *Harita Dergisi*, 159(1), 1-15, 2018.
- Sakellariou, M. G., Ferentinou, M. D., 2001. GIS-Based Estimation of Slope Stability. *Natural Hazards Review*, 2(1), 12-21, doi.org/10.1061/(ASCE)1527-6988(2001)2:1(12).
- Saltelli, A., Chan, K., Scott, E. M., 2000. *Sensitivity analysis*, UK, Wiley&Sons Ltd.
- Uğur, A., Özocaklı, D., 2018. Gıda Güvencesizliğinin Bazı Belirleyicileri (Kantil Regresyon Yöntemi Ve Sabit Etki Panel Yönteminin Karşılaştırılması). *Sosyoekonomi*.